GRAVITATIONAL WAVES IN THE POST-DETECTION ERA:

CHARACTERISING BINARY BLACK HOLE SIGNALS AND POPULATIONS

by Sebastian Michael Gaebel

A thesis submitted to the
University of Birmingham
for the degree of Doctor of Philosophy

School of Physics and Astronomy
and
Institute for Gravitational Wave Astronomy
College of Engineering and Physical Sciences
University of Birmingham

September 2019

Abstract

The detection of GW150914 has marked the start of the era of gravitational wave (GW) astronomy, and the number of detections available for astrophysical interpretation has been growing ever since. Similarly, this thesis progresses from considering individual signals to analysing the population of gravitational wave sources.

We explore how much we expect to be able to learn from a loud binary black hole merger like GW150914, were it to appear in the detector networks we currently have, as well as those which are being planned or under construction. Afterwards, we develop a model which allows us to obtain unbiased estimates of the rate with which the universe produces such signals, and the distribution of intrinsic gravitational wave source parameters such as the component masses, despite the fact that the observed distribution is inherently biased by the fact that source parameters affect the loudness of the emitted signal, and contaminated by terrestrial noise artefacts inherent to the detectors. We apply this method to a number of test cases.

Acknowledgements

I would like to thank all the different people who helped me on my journey through life so far, leading up to this moment. This work would not have been possible without your support. This especially includes my family – Michael, Petra, Annick, and Katharina Gaebel, as well as Eleonore Wagener and Eva Schumann – who were there for me throughout these last 29 years, and I cannot overstate how much you all mean to me. The last few months were particularly hard, but I made it through thanks to you.

On the academic side I would like to thank my supervisors – John Veitch, Ilya Mandel, Alberto Vecchio, and Patricia Schmidt – for their guidance, support, and feedback.

As for my fellow PhD students – Coenraad Neijssel, Daniel Töyrä, Sam Cooper, Alejandro Vigna-Gómez, Hanah Middleton, Anna Green, Janna Goldstein, Serena Vinciguerra, Siyuan Chen, Carl-Johan Haster, and all the ones I missed here – I can only say thank you for the last 4 years, and good luck going forward!

Last but not least, I would like to thank the other staff – especially Conor Mow-Lowry, Elinore Roebber, David Stops, Alberto Sesana, and Ian Stevens – for their help and interesting discussions throughout the last years.

My research has been supported by STFC grant ST/K005014/2.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

After decades of construction, research, and design efforts the LIGO and Virgo detectors have finally reached the sensitivity required to detect gravitational wave signals. Starting September 14, 2015 [1] we are able to detect some of the faint gravitational wave signals travelling the universe, as proven by the detection of GW150914, and use them to expand our knowledge of gravitational waves, general relativity, massive stars, and physics in general. Before that day gravitational waves were strongly suspected to exist, as there was strong indirect evidence published 30 years earlier by Weisberg et al. [2] observing the inspiral of the pulsar and neutron star system found by Hulse and Taylor [3] in 1981, but they had not been detected directly.

This work is split into three principal components which explore different, although related, aspects of gravitational wave astronomy.

Chapter 1 reviews the basics of gravitational wave emission (1.1, 1.2), detectors (1.3), and the tools required to extract astrophysical information from the output of gravitational wave detectors (1.4, 1.5, 1.6, 1.7). It concludes with an overview of selection bias in gravitational-wave astronomy (1.8).

Chapter 2 is adapted from a paper [4] written together with John Veitch, in which we explore the expected performance of future gravitational wave detector networks, using the very first detected GW signal (GW150914) as a reference point. To account for the addition of detectors and the fact that technical limitations prevent them from observing truly continuously we consider a number of different network configurations consisting of the three currently operational detectors (LIGO Hanford, LIGO Livingston, and

Virgo), as well as the two observatories under construction (KAGRA in Japan and LIGO India).

In Chapter 3 we consider the population of gravitational wave sources as a whole, and present a new model for characterising this population in the knowledge that the set of observed signals can (and, eventually, will) be contaminated by terrestrial noise and biased through selection effects. It is based on a paper [5] written in collaboration with John Veitch, Thomas Dent, and Will M. Farr.

We conclude in chapter 4.

1.1 General Relativity & Gravitational Waves

General relativity, as proposed by Einstein [6] is our current best theory of gravity. Over the course of the last century a number of experiments have been proposed and carried out, none of which has found any significant evidence against it [7, 8]. These experiments include, amongst others, the observation of light deflection [9], and the precession of Mercury's perihelion [6], and most recently the direct observation of the accretion disk of a supermassive black hole [10]. One of its fundamental predictions is the existence of gravitational waves [11] (GWs), fluctuations in the curvature of space-time which spread through the universe at the speed of light.

Gravitational waves are a small perturbation $h_{\mu\nu}$ of the curvature of space-time $g_{\mu\nu}$ which obey the wave equation

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} + \mathcal{O}(h^2) \tag{1.1}$$

$$\Box h_{\mu\nu} = 0 \tag{1.2}$$

where $h_{\mu\nu} \ll 1$. Equation 1.2 assumes us to have put $h_{\mu\nu}$ in transverse-traceless gauge, and for the gravitational wave to travel through flat, empty space. We know from conservation laws that mono- and dipoles cannot emit gravitational radiation [12], leaving the quadrupole as the leading order contributor to GW emission. This leads us to split gravitational waves into *plus* ("+") and *cross* ("×") polarised components.

The observable effects of GWs are best illustrated by considering a pair of test particles at rest. Once the wave reaches the particles, the proper distance changes, it oscillates around their initial separation with a dimen-

sionless amplitude of $(1 + h)$, i.e. $(1 + A_+)$ or $(1 + A_\times)$ along the axes as determined by the respective polarisation vectors, and transverse to the direction of propagation of the gravitational wave [12].

The luminosity L_{GW} of a gravitational wave emitter is [12]:

$$L_{\mathrm{GW}} = \frac{1}{5} < \left(\frac{\mathrm{d}^3\mathcal{I}}{\mathrm{d}t^3}\right)^2 > \tag{1.3}$$

$$\mathcal{I}_{jk} = \sum_n m_n \left(x_{nj}x_{nk} - \frac{1}{3}\delta_{jk}||\vec{x}_n||^2\right) \tag{1.4}$$

$$h_{jk}(t) = \frac{2}{d_{\mathrm{L}}}\frac{\mathrm{d}^2\mathcal{I}_{jk}}{\mathrm{d}t^2}(t - d_{\mathrm{L}}) \tag{1.5}$$

where $< \cdot >$ denotes the average over multiple wavelengths, $|| \cdot ||$ denotes the norm of a vector, d_{L} is the luminosity distance between the observer and the source, and $\mathcal{I}$ is the reduced quadrupole moment defined here as a sum over point masses. Equations 1.3 and 1.4 reveal why compact binary coalescences are prime sources of GW signals. The proportionality relations in these equations are:

$$\dddot{\mathcal{I}} \propto \frac{MR^2}{T^3}, \tag{1.6}$$

$$L \propto (\dddot{\mathcal{I}})^2 \propto \frac{M^2R^4}{T^6} \propto \frac{M^5}{R^5}, \tag{1.7}$$

where M, R, and T are the mass, size, and dynamic time-scale of the system. We used Kepler's third law to eliminate T in favour of M and R. These expressions show that the gravitational wave emission from binary systems is at its largest for sources which concentrate a large mass in a small area at high velocities, which is the case in binaries composed of black holes (BH) or neutron stars (NS).

A more detailed calculation (see Misner et al. [12]) yields

$$\begin{aligned} L_{\mathrm{GW}} &= \frac{32}{5}\frac{\mu^2M^3}{r^5}f(\epsilon)L_0 \\ &= \frac{32}{5}\frac{\mathcal{M}^5}{r^5}f(\epsilon)L_0 \end{aligned} \tag{1.8}$$

where we define:

- m_1 and m_2 are the component masses ($m_1 \geq m_2$),
- $M = m_1 + m_2$, the total mass,

- $\mu = \frac{m_1 m_2}{M}$, the symmetric mass ratio,
- $\mathcal{M} = \mu^{3/5} M^{2/5}$, the chirp mass,
- r is the separation of the binary components,
- ϵ, is the eccentricity,
- $f(\epsilon) = \left[1 + \frac{73}{24}\epsilon^2 + \frac{37}{96}\epsilon^4\right] \left(1 - \epsilon^2\right)^{-7/2}$,
- $L_0 = c^5 G^{-1} \approx 10^{53}$W.

Since isolated black holes do not emit significant amounts of energy [13], the energy loss of a binary black hole is dominated by the emitted GWs which cause the orbital separation to decrease and the orbital (and thereby gravitational wave) frequency to increase. Using Equation 1.3, $E = -\frac{1}{2}\mu M r^{-1}$, and assuming the eccentricity to be $\epsilon = 0$, we can derive a simple expression for the evolution of the orbit due to GW emission:

$$\frac{1}{2}\frac{\mu M}{r^2}\frac{\mathrm{d}r}{\mathrm{d}t} = \frac{\mathrm{d}E}{\mathrm{d}t} = -L_{\mathrm{GW}} = -\frac{32}{5}\frac{\mu^2 M^3}{r^5}, \tag{1.9}$$

$$\implies r = r_0 \left(1 - \frac{t}{t_c}\right)^{1/4}, \tag{1.10}$$

$$t_c = \frac{5}{256}\frac{r_0^4}{\mu M^2} \tag{1.11}$$

where t_c is the time-scale on which a binary spirals-in to merge, and r_0 is the initial orbital separation.

Naturally this approximation breaks down when the components of the binary grow too close, at which point one must either add additional post-Newtonian (pN) corrections or transition to solving the field equations directly.

1.2 Gravitational Wave Signals from Compact Binaries

The evolution of a binary black hole due to the emission of gravitational waves can broadly be split into three phases: Inspiral, merger, and ringdown.

During the inspiral phase the binary components are at such large separations that quasi-Newtonian solutions as used to derive Eq. 1.10 with pN corrections are sufficient to describe the evolution of the binary.

Eventually, once a time t_c has passed, the components grow close enough for these approximations to fail. The binary's orbit will continue to shrink at

an accelerating pace under the increasing emission of GWs until the event-horizons of the two components overlap, at which point the two components merge into a single compact object. During this phase space-time between and around the two components is highly dynamic and determining the behaviour of the system typically requires numerical solutions to the field-equations (*numerical relativity*, NR) [14, 15, 16].

Once the binary has merged, the system will emit some last GWs as the merger product settles down into its final state. For a black hole, this ringdown phase is characterised by the quasi-normal oscillation of the event horizon during which an exponentially decaying, sinusoidal GW signal is emitted [17].

While NR simulations are possible, they are computationally intensive and time consuming. Searching for gravitational wave signals or analysing them in detail typically requires the generation of thousands to millions of simulated signals [18]. These waveforms cannot be obtained at a reasonable computational cost if NR simulations are performed for each individual waveform, even if combined with analytic approximations for the inspiral or merger components. This issue is tackled by the development of waveform approximants, functions which attempt to accurately capture all important features of the predicted signals, but at much lower computational cost. The two main approaches are the EOB and IMRPhenom families of approximants.

The EOB approach [17, 19, 20] uses the **E**ffective **O**ne **B**ody formalism to approximately solve the inspiral component and combines it with a fit to NR results in the merger and ringdown regime. The IMRPhenom (**I**nspiral-**M**erger-**R**ingdown, **Phenom**enological) family [21, 22, 23] takes a slightly different approach, which is to build a reduced order surrogate model based on fits to pN results and NR simulations.

These approximants generally exclude many higher order features of the signals, though development is ongoing to include more detail (e.g. precession or higher modes) and improve their general accuracy [24].

1.3 Detecting Gravitational Waves

To be able to detect a passing gravitational wave we need to recall their effect on an observer. Consider a pair of test particles A and B at rest relative to

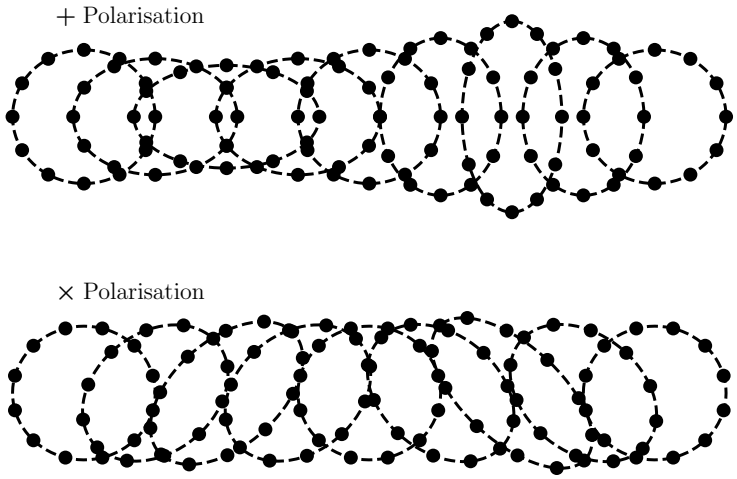

Figure 1.1: A ring of test particles at rest under the influence of a passing gravitational wave. The circular configurations at the very right, left, and in the centre, show the ring while the gravitational wave amplitude is zero. Bottom panels: effect of the "cross" polarisation component. Top panels: effect of the "plus" polarisation component. The gravitational wave amplitude of $h \approx 0.2$ is greatly exaggerated for illustration purposes.

each other. Using the transverse nature of gravitational radiation, Misner et al. [12] show that in the presence of a gravitational wave $h(t)$ propagating along $\vec{e}_h$ the proper separation $\vec{n}_{AB}$ becomes

$$\vec{n}_{AB} \mapsto \begin{cases} \vec{n}_{AB}(1 + h(t)) & \text{if } \vec{n}_{AB} \perp \vec{e}_h, \\ \vec{n}_{AB} & \text{if } \vec{n}_{AB} || \vec{e}_h, \end{cases} \tag{1.12}$$

$$\frac{\Delta||\vec{n}_{AB}||}{||\vec{n}_{AB}||} = h(t)\sin(\varphi), \tag{1.13}$$

where φ is the angle between $\vec{n}_{AB}$ and $\vec{e}_h$. If $\vec{n}_{AB}$ has both parallel and orthogonal components, these transformations apply component-wise. Recalling that gravitational radiation is composed of $+$ and $\times$ components, this results in a behaviour best illustrated by a ring of test particles, as shown in Figure 1.1.

As a result, any device capable of measuring proper distances between objects is a potential GW detector. Equation 1.13 shows that the dimension-

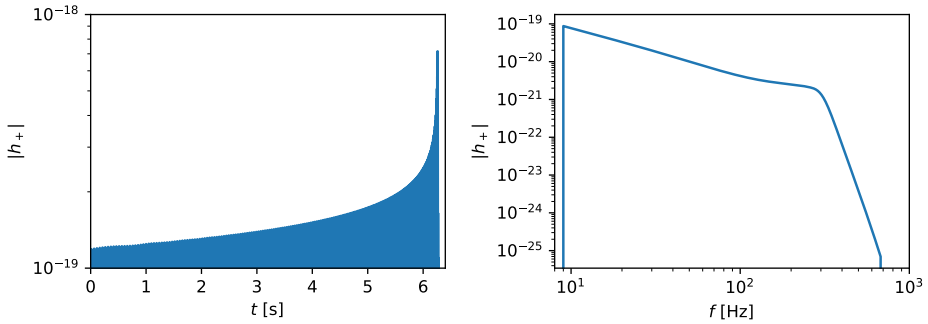

Figure 1.2: An example gravitational wave, shown both in frequency-domain (right) and time-domain (left). This example used the IMRPhenomPv2 waveform approximant for a system of two 30 $M_\odot$non-spinning black holes at a distance of 1 Mpc. The y-axis shows the dimensionless strain amplitude of the "+" polarisation.

less GW strain $h(t)$ directly translates into the fractional change in proper distance, which is proportional to the signal in any detector based on this effect. In practice the extremely small amplitudes of these signals, the maximum gravitational wave strain amplitude observable during the merger of a 30 & 30 $M_\odot$binary in Andromeda is $< 10^{-18}$ (see Figure 1.2), makes designing such a detector rather challenging.

The currently favoured detector design is the laser-interferometer. The LIGO and Virgo detectors are based on the Michelson interferometer with numerous changes, additions and improvements [25]. The basic layout of such a detector is shown in Figure 1.3. This type of detector is sensitive to the change in differential arm lengths, i.e. $d(t) \propto |L_x(t) - L_y(t)|$ where d is the "data" produced by the detector and L_x and L_y are the lengths of arm X and Y respectively. The arm length in this case is measured between the two test mass mirrors in each arm, as the effective length of the cavity dwarfs all other distances travelled by the beam.

Since any real detector is affected by noise, the detector output $d(t)$ can be written as the sum of the gravitational wave signal $h(t)$ and some noise $n(t)$

$$d(t) = h(t) + n(t). \tag{1.14}$$

The signal is proportional to the projection of the gravitational wave onto

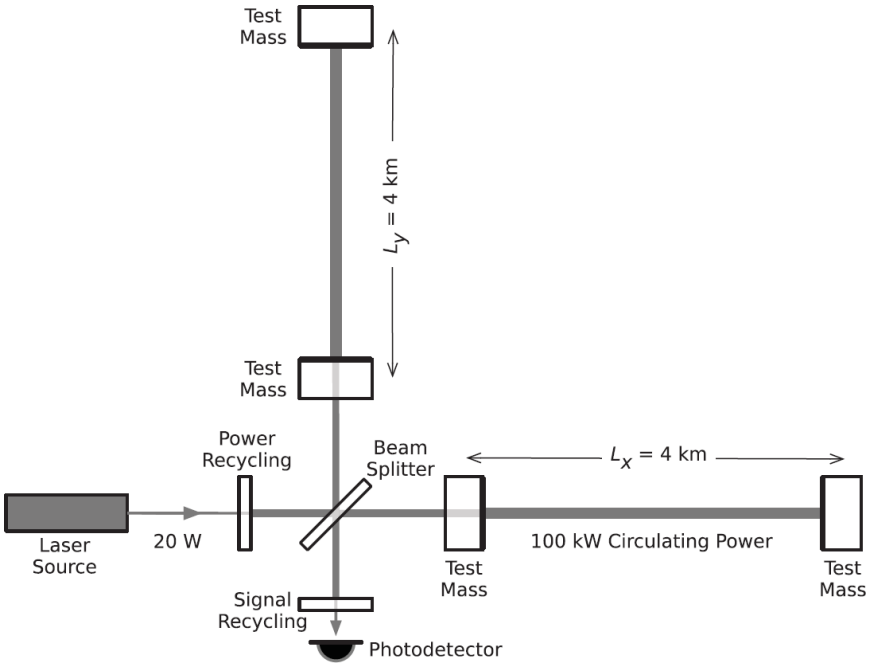

Figure 1.3: Simplified layout of a laser-interferometer gravitational wave antenna. The laser enters the device from the laser source on the left and is split after passing the power recycling mirror. Each half of the original beam propagates through the test mass of either the x or y arm and spends some time circulating in the arm cavity, before leaving through the test mass near the beam splitter. The light is then recombined by the beam splitter and leaves (partially) through the signal recycling mirror to be detected by the photo-detector. This illustration has been adapted from Figure 3 of Abbott et al. [1] with minor edits to isolate the part of the figure relevant in this context.

the arms of the detector. Formally we can write

$$h(t) = F_+ h_+(t) + F_\times h_\times(t), \tag{1.15}$$

where F_+ and $F_\times$ are the projection operators which describe the effect of the respective gravitational wave polarisations on the detector. These projection operators are generally time dependent due to the rotation of the earth, but the time dependence is negligible for any given CBC event since the GW signals considered throughout this text are of order seconds to minutes (see Equation 1.11).

Using the Newtonian approximation from Equation 1.10 the inspiral signal as observed by a LIGO-like detector can be written as [26]

$$h(t) = \sqrt{\frac{384\pi^{4/3}}{5}} \frac{Q(\theta, \varphi, \psi, \iota)\mu M}{r(t) d_\mathrm{L}} \cos\left(\phi(t)\right) \tag{1.16}$$

$$r(t) = \left(\frac{256}{5}\mu M^2 (t_c - t)\right)^{1/4} \tag{1.17}$$

$$\phi(t) = \phi_c - 2\left(\frac{t_c - t}{5\mathcal{M}}\right)^{5/8} \tag{1.18}$$

where d_L is the luminosity distance between source and observer, $\phi(t)$ is the phase of the GW signal, and $Q(\theta, \varphi, \psi, \iota)$ encapsulates all effects caused by the relative orientations of binary and detector. In practice, the more sophisticated waveform approximants mentioned in Section 1.2 are used instead.

The noise component $n(t)$ of Equation 1.14 is generally well approximated as a draw from a coloured, stationary, and Gaussian noise with some noise spectrum $S_n(f)$. This spectrum depends on the exact configuration of the detector, and therefore changes as improvements and modifications are applied. During observation periods such changes are kept to the absolute minimum to ensure stable operating conditions, though these efforts may be undermined by environmental effects, e.g. storms or earthquakes. On average the duty cycle of the two LIGO detectors during the first two observing periods (O1 & O2) was $\approx 50\%$ [28]. Figure 1.4 shows the measured noise spectra of the two LIGO observatories during O1 and O2 each, as well as the planned noise spectrum for their final configuration. It also shows the spectrum of an example stellar mass CBC signal (a BBH with

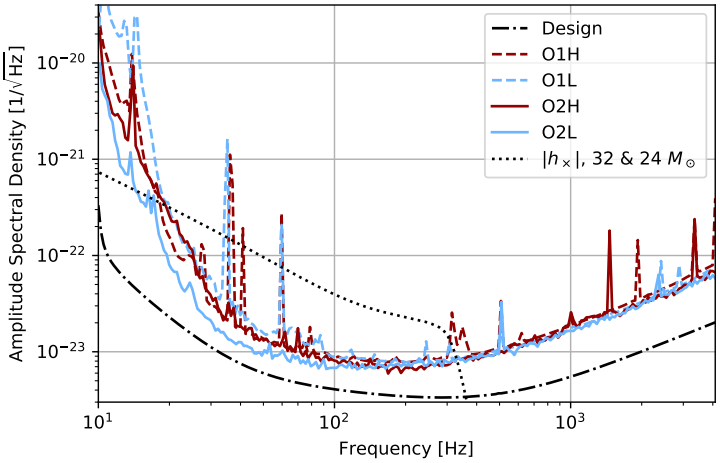

Figure 1.4: Noise levels measured during the first two observing runs and expected noise levels at Design sensitivity. The first observing run [27] is plotted using dashed lines, the noise levels during the second observing run [28] are given using solid lines. The Hanford and Livingston LIGO detectors are shown in red and blue, respectively. The solid black line shows the sensitivity LIGO was designed to target [29], though further upgrades beyond this being considered. The dashed black line is the frequency domain waveform of a 32 & 24 $M_\odot$, non-spinning, face-on binary black hole coalescence placed directly over the detector at a distance of 100 Mpc.

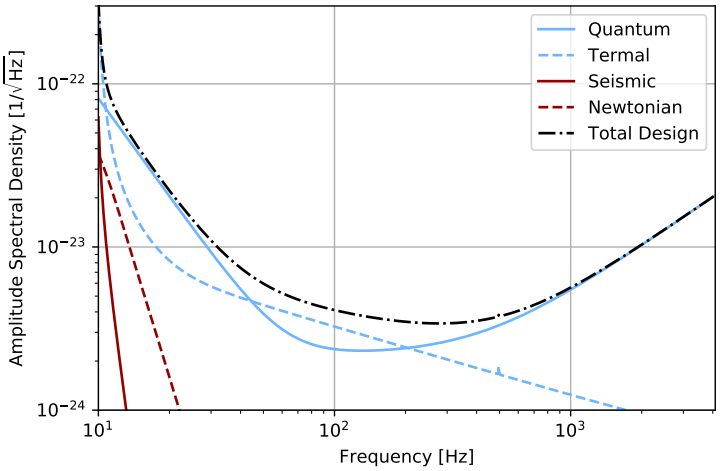

Figure 1.5: Decomposition of the Design-sensitivity noise ASD into the different noise sources. Shown are *i)* Total (black, solid), *ii)* Quantum (blue, solid), *iii)* Thermal (blue, dashed), *iv)* Seismic (red, solid), and *v)* Newtonian (red, dashed). Noise simulated using `pygwinc` [30].

non-spinning components, 32 & 24 $M_\odot$ each, simulated using the implementation in `lalsimulation` [31] of the IMRPhenomPv2 approximant).

The shape of the detectors' noise spectra is caused by the frequency dependencies of the different noise sources. The most important contributors to the total noise budget are

Quantum: quantum uncertainties of e.g. photon count in the laser (shot noise) and the fluctuations in the photon pressure on the mirrors,

Thermal: caused by Brownian motion of the suspensions and the surfaces of mirrors,

Seismic: ground motion due to e.g. earthquakes or pedestrians,

Newtonian: changes in the local gravitational field due to mass moving near the detector.

Their frequency dependencies are illustrated in Figure 1.5. To suppress seismic noise the end mirrors are suspended by silicate fibres in multi-stage pendulum, fine-tuned to filter out the frequency bands responsible for most of the noise transmitted to the test-masses.

1.4 Finding Gravitational Wave Signals

When searching for signals in some stream of data, it is essential to be able to quantify how well any given simulated signal matches the observed data. The quantity used to do so is the signal-to-noise ratio (SNR, ρ), defined as

$$\mathrm{SNR} = \rho = \sqrt{4\mathcal{R}\left[\int_0^\infty \mathrm{d}f \frac{\tilde{h}^*(f)\tilde{h}(f)}{S_n(f)}\right]} \tag{1.19}$$

where $\tilde{h}(f)$ is the complex frequency domain GW signal. Using the Newtonian approximation and Equation 1.16 we obtain [32]

$$\rho \propto \frac{\mathcal{M}^{5/6}}{d_\mathrm{L}} \sqrt{\frac{5}{96}\pi^{-4/3} \int_0^\infty \mathrm{d}f [f^{7/3} S_n(f)]^{-1}}. \tag{1.20}$$

From Equation 1.20 we can directly derive the distribution of SNR val-

ues:

$$p(\rho, \mathcal{M}) \propto \int_0^\infty \mathrm{d}d_\mathrm{L}\; 4\pi d_\mathrm{L}^2 p(\rho, \mathcal{M}|d_\mathrm{L}) \tag{1.21}$$

$$\propto \int_0^\infty \mathrm{d}\hat{\rho}\; 4\pi \hat{\rho}^{-2} \mathcal{M}^{10/6} \times \mathcal{M}^{5/6}\rho^{-2} \times p(\rho, \mathcal{M}|\hat{\rho}) \tag{1.22}$$

$$\propto \rho^{-4}\mathcal{M}^{15/6} \tag{1.23}$$

$$\text{since } \frac{\mathrm{d}d_\mathrm{L}}{\mathrm{d}\rho} \propto \mathcal{M}^{5/6}\rho^{-2} \tag{1.24}$$

This assumes GW sources to be uniformly distributed in luminosity volume ($p(d_\mathrm{L}) \propto d_\mathrm{L}^2$) and ignores the cosmological redshifting of the signal. At this point, we can also note that the observed $\mathcal{M}$-distribution is biased by a factor of $\approx \mathcal{M}^{2.5}$. We will revisit this effect later in Section 1.8 and Chapter 3.

If sources are detected to distances large enough for cosmological effects to be significant, the previous relation only holds for small distances. Figure 1.6 shows the analytic curves for the distribution of optimally oriented sources in distance as well as two histograms showing the results of simulating a large number of sources distributed either uniformly in luminosity volume with a sharp cut-off at 14 Gpc (luminosity), or uniformly in co-moving volume with a cut-off at 5 Gpc (co-moving, redshift $z \approx 1.8$). Deviations from the analytic curve are expected as those do not account for horizons or SNR variations due to source and detector orientation.

This information is used to search through the data taken by the detectors and find any possible gravitational wave signals contained. We know how to simulate a GW signal, how it appears in the detector output, we have a detection statistic (e.g. SNR), and we even know the expected distribution of SNR values. For computational efficiency search algorithms typically use a template bank, a precomputed set of typically 10^5–10^6 simulated GW signals constructed carefully to recover as many signals as possible, without wasting computational resources. For each template in the bank the detection statistic is computed at each point of the available data.

The detection statistic by itself is not sufficient to efficiently isolate (likely to be) real signals. For this, we need quantify the distribution of detection statistic values caused by noise alone, and compare it to the expected foreground distribution we are able to predict, as done in Equations 1.21–1.24. While this may be easy to accomplish for most detectors or telescopes by

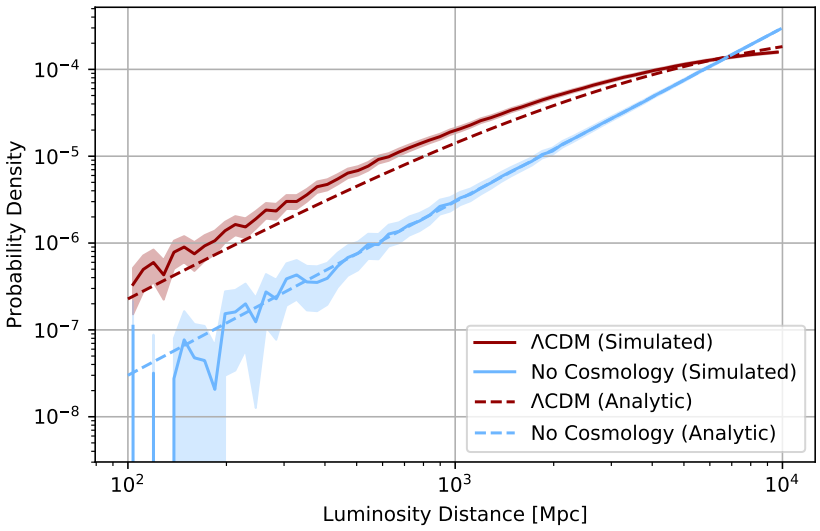

Figure 1.6: Distribution of source-observer-separations. The dashed lines show the analytic distributions assuming no horizon. The solid lines show the result of a simulated distribution of signals. The simulations take into account the random orientation of source and observer, existence of horizons (chosen to be at $d_{\mathrm{L}} = 14\mathrm{Gpc}$ if cosmology is ignored, and $d_c = 5\mathrm{Gpc}$, $z \approx 1.8$). The simulation which includes cosmological effects also includes the changed sensitivity of the detectors to redshifted GW signals. The light regions around indicate the estimated uncertainty of the histogram due to Poisson uncertainty in the number of points in a given bin. The cosmology is simulated using a flat ΛCDM model with $H_0 = 67.7\frac{\mathrm{km}}{\mathrm{sMpc}}$, $\Omega_{\mathrm{M}} = 0.307$, and $\Omega_\Lambda = 0.693$ as given by Planck Collaboration et al. [33] using `astropy` [34, 35]. Most sources are located at very large distances, therefore their signals are very weak and difficult to observe. Nearby sources often do not require us to consider cosmological effects, which become important once the luminosity distance reaches $\approx 1\mathrm{Gpc}$.

simply taking data while not exposing them to the a source (as done in Astronomy routinely through dark-frame subtraction), gravitational wave detectors can not be shielded from gravitational radiation. Fortunately though, gravitational wave detectors can not be shielded from gravitational radiation, even by putting massive objects like planets between source and observer. This means we expect the signals detected by two separate detectors to be indistinguishable after taking into account wave propagation effects such as the difference in distance to the source (at 1 kHz the wavelength $\lambda \approx 300$km), the difference in how the detectors are spatially oriented relative to the source, or different detector noise spectra. Since there are currently 3 GW observatories in operation, we exclusively search data taken simultaneously at 2+ sites so we are able to reject signals not appearing in all detectors.

One approach to estimating the background is use the same idea in reverse by sliding the data from the different sites against one-another. If the offset is larger than the light travel time between the detectors, we can be certain that any coincident signal must be caused by noise artefacts and estimate the detection statistic distribution of the background. A second advantage of the technique is that the effective amount of time available to analyses is much larger, as the number of possible time-shifts grows linearly with the length of the data segment, i.e. a 10s data segment has an effective duration of $10\text{s} \times \frac{10\text{s}}{2\times 10\text{ms}} = 500\text{s}$ assuming 10ms light travel time and safety factor of 2, while $100\text{s} \mapsto 50,000\text{s}$. Using this technique we find the background distribution $p(\rho|\text{BG})$ to roughly follow a power-law with exponent -50: $p(\rho|\text{BG}) \propto \rho^{-50}$. These estimates should only be taken as rough estimates though, the distribution and exponent change with noise characteristics and detection pipelines.

Once the statistical properties of both foreground and background are known, the ratio between foreground and background probabilities allows us to label any event with a probability of being of astrophysical origin (as opposed to being caused by terrestrial noise artefacts)

$$p_{\text{astro}}(\rho) = \frac{p(\rho|\text{FG})}{p(\rho|\text{FG}) + p(\rho|\text{BG})}. \tag{1.25}$$

This relation ignores any pre-existing knowledge we have about the properties of astrophysical GW source population or effects such as dependencies

of the distributions on source parameters. A more detailed discussion and implementation of an improved model is given in Chapter 3.

The two main detection pipelines used by the LIGO and Virgo detectors during the first two observing runs are `pycbc` [36], which uses the time-slide method outlined above to characterise the background, and `gstlal` [37], which uses non-coincident single-detector triggers instead.

An additional benefit of using multiple (N) GW antennas is that the effective SNR of the observed signal is larger by a factor of $\sqrt{N}$ since the noise realisations are independent and will, on average, cancel. It also helps in determining the origin of a given signal, since an individual antenna observes $\approx 50\%$ of the sky with little ability to localise source further, while 2+ antennas can triangulate the source location by comparing arrival times and detector sensitivity patterns [38]. As a consequence most detections from O1 & O2 are localised to a ring on the sky due to being observed by two detectors, while three-detector observations like GW170814 [39] are constrained to a small circular region. These effects are especially important as we continue to improve the sensitivity of the existing detectors, as well as to build new ones, and are discussed in greater detail in Chapter 2.

1.5 Bayesian Statistics

Since the next section will discuss the detailed analysis of individual GW signals to determine its source properties, this section will briefly introduce Bayesian statistics [40] and why we choose to base our analysis on it.

Unsurprisingly, Bayesian statistics is centred around Bayes' theorem

$$p(A|B,M) = \frac{p(B|A,M)p(A|M)}{p(B|M)} \tag{1.26}$$

where A is the parameter we wish to infer, B is the observation or experiment we use for our inference, and M is the model which describes how A and B are related. The individual components are:

$\boldsymbol{p(A|B,M)}$ is the *posterior* probability of A given B.

$\boldsymbol{p(B|A,M)}$ is the *likelihood* of observing B if given A.

$\boldsymbol{p(A|M)}$ is the *prior* probability of A, which expresses our knowledge prior to the experiment.

$\boldsymbol{p(B|M)}$ is the *evidence* or *marginal likelihood*.

The posterior distribution quantifies our belief in how likely a parameter (or set thereof) is to have some specific value, given the observational data, model, and the knowledge we had prior to the experiment. The evidence term is often ignored as it can be hard to compute, and is constant for any given set of observations B. If only the posterior distribution is of interest, the evidence is simply a normalisation constant and does not affect our inference on A. It does, however, have an important application when comparing how well different models fit a given dataset (hence the name *evidence*), as it describes how likely it is to observe the data without any assumptions on the model parameters A (i.e. the likelihood after marginalising over A). This is accomplished by computing the Bayes factor $\frac{p(B|M_1)}{p(B|M_2)}$, which quantifies how much more likely model M_1 is to produce the observed data B than model M_2.

This formalism can be extended to either include a mixture of different sub-populations, or contain a hierarchy of models. In case of a hierarchical model, the posterior from one layer acts as the prior for the next lower layer. A simple example involving compact binary gravitational wave sources might be the case where, assuming all gravitational wave signals are generated by such binary coalescences, each event has a set of true physical parameters γ (e.g. component masses or spins) drawn randomly from a population of sources (model M) characterised by a set of parameters θ, and whose parameters are then estimated given the strain data $\vec{d}$ and some waveform approximant (model W):

$$p(\theta, \gamma|\vec{d}, M, W) \propto p(\theta|M_1)\ p(\gamma|\theta, M_1)\ p(\vec{d}|\gamma, W), \tag{1.27}$$

$$p(\theta|\vec{d}, M, W) \propto \int \mathrm{d}\gamma\ p(\theta|M)\ p(\gamma|\theta, M)\ p(\vec{d}|\gamma, W), \tag{1.28}$$

where we have marginalised over the event parameters γ in Equation 1.28 to obtain the likelihood for the population parameters θ only, while the expression Equation 1.27 provides a joint posterior on θ and γ. This formalism can easily be generalised to multiple events by either applying the model multiple times using the posterior of the previous iteration as the new prior, or introducing a separate γ for each event, which is simplified if those events are mutually independent.

When the goal is to obtain estimates of the parameter for an individual

source the likelihood is similarly

$$p(\gamma|d, M) \propto p(d|\gamma, M)p(\gamma|M), \tag{1.29}$$

with the difference that the prior is fixed for any given signal. As it is usually possible to incorporate changes to the prior after the likelihood or posterior distribution have been calculated, it is often useful to analyse all events separately with some fixed and uninformative prior using Eq. 1.29, and use those results in the hierarchical analysis to determine the population using Eq. 1.27.

A mixture model with two mutually exclusive sub-populations M_1 and M_2 is trivially derived:

$$p(A|B, M_1 \vee M_2) = p(A|B, M_1)p(M_1) + p(A|B, M_2)p(M_2), \tag{1.30}$$

where Bayes' theorem can now be applied separately on both components. Here A is the inferred parameter, M_1 and M_2 are two models for how A affects B. This yields:

$$\begin{aligned} p(A|B, M_1 \vee M_2) =& \frac{p(A|M_1)p(B|A, M_1)}{p(B|M_1)}p(M_1) \\ &+ \frac{p(A|M_2)p(B|A, M_2)}{p(B|M_2)}p(M_2). \end{aligned} \tag{1.31}$$

Note that we can not drop the evidence terms unless $p(B|M_1) = p(B|M_2)$, as this would affect the relative contributions of the two models to the posterior probability density.

Hierarchical and mixture models can be combined arbitrarily, which can quickly lead to complex expressions and non-trivial dependencies between the different components.

1.6 Characterising Gravitational Wave Signals

Once time segments likely to contain astrophysical gravitational wave signals are identified, we would like to characterise the exact nature of the source producing any given signal, as well as the associated uncertainties. The key

quantity for this process is the likelihood:

$$p(\vec{\gamma}|d) = \mathcal{L}(\vec{\gamma}) \propto \prod_{i=0}^{N-1} \exp\left[-\frac{1}{2}\left(\frac{d(t_i) - h(t_i, \vec{\gamma})}{\sigma_n}\right)^2\right] \tag{1.32}$$

$$= \exp\left[-\frac{1}{2}\sum_{i=0}^{N-1}\left(\frac{d(t_i) - h(t_i, \vec{\gamma})}{\sigma_n}\right)^2\right] \tag{1.33}$$

which is the answer to the question "how likely is the data $d(t)$ assuming a signal $h(t, \vec{\gamma})$ is present" or equivalently "how likely are we to observe the specific noise realisation $n(t)$", where $d(t) = h(t, \vec{\gamma}) + n(t)$ as in Equation 1.14, and $\vec{\gamma}$ is the set of source parameters. Here we assume the data $d(t)$ is recorded at N discrete times t_i and a constant, white, Gaussian noise amplitude σ_n. For a real noise spectrum $S_n(f)$ we shift the analysis into the frequency domain, where we obtain [41]:

$$\mathcal{L}(\vec{\gamma}) \propto \exp\left[-\frac{1}{2} < d(t) - h(t, \vec{\gamma}) | d(t) - h(t, \vec{\gamma}) >\right], \tag{1.34}$$

where $< \cdot >$ is the noise-weighted inner product:

$$< m(t)|n(t) > = 4\,\mathcal{R}\int_{-\infty}^{\infty} \mathrm{d}f\ \frac{\tilde{m}^*(f)\ \tilde{n}(f)}{S_n(f)}, \tag{1.35}$$

and $\tilde{x}(f)$ is the (complex valued) Fourier transform of $x(t)$.

A simple but reliable estimate for the source parameters can be obtained by finding the values of $\vec{\gamma}$ which maximise $\mathcal{L}$ or, equivalently, $\log \mathcal{L}$, i.e. the point of the most likely noise-residual. A more detailed analysis uses the Bayesian framework given in the previous Section 1.5 to produce a continuous posterior distribution for the inferred parameters $\vec{\gamma}$ using the same likelihood function. This allows us to quantify the uncertainties in the inference and to identify other interesting features which would be overlooked by summarising the result via a point estimate, e.g. multi-modalities or correlations.

1.7 Exploring Parameter Space

Independent of how the posterior $p(A|B, M)$ is derived, extracting useful information requires us to explore its values across the space of allowed pa-

rameters, as defined by the prior $p(A|M)$. If A has a low dimension, this can easily be accomplished by evaluating the posterior at all points on a regular grid. Unfortunately the number of evaluations required grows exponentially as the number of dimensions increases and quickly becomes infeasible. The most commonly used alternative is to use Monte Carlo (MC) techniques [40] to draw samples from the high-dimensional posterior, where the density of samples is proportional to the posterior probability density. Using samples drastically reduces the amount of evaluations and storage needed since the samples naturally focus the used resources on the regions of high interest, i.e. regions with high probability density. This set of samples can, in turn, be used to compute all relevant quantities derived from the posterior distribution, and due to the central limit theorem these estimates converge to the true value as $N^{-1/2}$, where N is the number of statistically independent posterior samples. It is especially simple to marginalise over unneeded parameters, as all one need to do is remove or ignore the marginalised components of each sample.

The two methods used in the later chapters also represent two of the main schools of Monte Carlo sampling techniques: MCMC (Markov Chain Monte Carlo, introduced by Metropolis et al. [42] with improvements by Hastings [43]) and nested sampling (pioneered by Skilling [44][45]) which has the benefit of directly providing the evidence as well as posterior samples. Within the umbrella of Monte Carlo sampling there exists a wealth of different approaches and methods which aim to satisfy some combination of requirements on simplicity, ease of parallelisation, speed of convergence, and avoiding the need for tuning by humans. In the later chapters, we use a Ensemble MCMC technique by Goodman and Weare [46], implemented by Foreman-Mackey et al. [47], as well as the custom nested sampler in `lalinference` for analysing GW strain data.

MCMC techniques generally work by distributing a number of "walkers" throughout the prior volume and then have them perform a random walk. The key ingredient is that the steps of the random walk are proposed (through various different methods, e.g. fixed distributions or based on the positions of other walkers) and accepted with a probability determined by the relative probability densities. This probabilistic acceptance ensures that the density of points accumulated throughout the random walks is, eventually, proportional to the probability density in any given region.

In nested sampling the core idea is to have points randomly explore the prior range, as opposed to the posterior for MCMC, until they arrive ("die") at a location which passes an ever increasing threshold of posterior probability density. This effectively draws shrinking contours around the most likely regions of parameter space, and the dead points accumulated throughout can be transformed into posterior samples, similar to those produced by MCMC.

In practice the key difference between MCMC and nested sampling is the fact that nested sampling directly computes the evidence (it is, in fact, a technique for computing evidences, the posterior samples were originally just a by-product), while MCMC techniques produce useable intermediate results and can be run for arbitrary amounts of time to produce as many or few samples as required.

Independent of how the posterior samples are obtained, their key defining feature is that each sample represents the same amount of posterior probability. As the average distance between samples is smaller in regions of high probability density, the product between probability density and volume remains approximately constant.

In addition to being a more efficient representation of a probability density function (pdf), using samples to compute statistical properties of a function is also computationally much cheaper. Imagine a function f from which we have drawn N samples x_i. To compute e.g. the mean of f, we can simply compute

$$< f > \approx \frac{1}{N} \sum_{i=0}^{N-1} \vec{x}_i. \tag{1.36}$$

Similarly, performing marginalisation is trivial for a set of samples as all it takes is to ignore the existence of (or delete) the relevant dimension of each sample $\vec{x}_i$.

1.8 Selection Bias

We saw in section 1.4 that the amplitude and frequency evolution, and therefore SNR, of the GW signal strongly depends on the source parameters. A key consequence is that signals from certain parts of the parameters space are more likely to be detectable, e.g. a signal with higher mass will produce a higher amplitude signal according to equation 1.16, all other pa-

rameters being equal. This introduces a bias in the observed distribution of sources which must be taken into account when attempting to characterise the source population. From equation 1.23 we can predict that the chirp mass distribution of the observed population will be biased by a factor of $\approx \mathcal{M}^{2.5}$, though this ceases to hold for high mass CBC events in which the inspiral contributes less to the total detectable signal.

In practice, the selection effects are generally treated numerically, an analytic analysis becomes impractical as higher order effects and less impactful source parameters are considered. Refinements include the frequency dependence of the detector sensitivity, the inclusion of cosmological effects, and the use of more accurate waveform models. Therefore the typical approach to this is to *inject* known signals into (a copy of) the data, re-apply the search algorithms, and cross-reference found signals with known injections.

Similarly, these effects must be taken into account when simulating the formation and origin of GW sources. In this case, the intrinsic source parameters are provided as the output of the simulation, while most extrinsic parameters such as the relative orientation of source and detector are the result of a simple stochastic process. Often, the selection effects can be approximated numerically using the method described by Finn and Chernoff [32], in which only a single detector is used. This approach is especially applicable to observations made with the two LIGO detectors only, as they are aligned in such a way that the sensitivity patterns nearly match. In this case the SNR threshold, the minimum SNR above which a segment is considered to contain a likely signal, is adjusted by a factor $\sqrt{\#\text{Detectors}}$. It also allows for any waveform to be re-used at arbitrary distances and angles since those can easily be applied in a step separate from the (relatively) expensive evaluation of the GW waveform approximant. This method has been implemented by the author to be used by the COMPAS team [48, 49] to facilitate a comparison of the simulated CBC populations to those observed by LIGO & Virgo.

Chapter 2

How would GW150914 look with future gravitational wave detector networks?

Adapted from [4] with minor changes and updates. My contribution to this work was to a) *run the simulations,* b) *generate plots,* c) *write the majority of the paper. This paper was published in* Classical and Quantum Gravity *in September 2017.*

2.1 Introduction

The first gravitational wave signal GW150914, from a black hole binary merger, was observed by the two Advanced LIGO (aLIGO, [50]) detectors in Hanford and Livingston [1]. The masses of the two black holes were inferred to be $36.2^{+5.2}_{-3.8}\,\mathrm{M}_{\odot}$and $29.1^{+3.7}_{-4.4}\,\mathrm{M}_{\odot}$in their rest frame, forming a merger product of mass $62.3^{+3.7}_{-3.1}\,\mathrm{M}_{\odot}$ [51, 52]. The sky localisation of this event was poorly constrained as it is largely determined through triangulation using the difference in arrival time at the active detectors, and with only two operating aLIGO detectors the position was resolved to an annulus within a ring of constant time delay between the two sites [51]. However the Advanced Virgo (AdVirgo, [53]) detector has now joined the network, with KAGRA [54] and LIGO-India [55] expected to follow over the next years [29]. This raises the question as to how well those future networks can be expected to localize an event like GW150914, and how well its parameters could be measured with

the upcoming second generation detector networks. The subsequent detection of GW151226 and GW151012 [18, 52], as well as the 7 binary black hole signals from the second observing run [28] provide evidence for a population of massive black hole binaries, which are likely to produce multiple further detections in the future [56, 52].

Projections for future sensitivity improvements and network configurations are given in [29], which also studies the sky location performance. However this study, in common with the majority of previous works [57, 58, 59, 60, 61], considers only the binary neutron star case. Expectations for localisation of generic systems were given in [38, 62] using geometric arguments, which are a useful guide for qualitative interpretation of actual simulations in the 3+ detector case. However, [63, 57] indicate quantitative differences between such arguments and full Bayesian parameter estimation results, and qualitative differences in the two-detector network from the availability of amplitude measurements. [64] studied the parameter estimation expectations for generic systems from a heavy BBH population, while focusing mainly on mass and spin measurements. Most of the results are obtained using a network of one AdVirgo and two aLIGO detectors, although the five-detector network including LIGO-India and KAGRA was considered in an appendix but without comparing identical events. [65] studied sky localisation for short transient signals, using generic burst algorithms, however these can be systematically different from sky localisation which uses a compact binary signal model [66].

In this work we address the question of localisation and parameter estimation for massive BH binaries from a different angle. Using GW150914 as a template, we perform a set of simulations based on an evolving network configuration, keeping the injected signals the same. This allows us to study the improvements in parameter estimation and localisation systematically, using the Hanford-Livingston (HL) network as a reference, and studying the separate improvements produced by the expansion of the detector network and the general increase in sensitivity of these detectors.

We considered a variety of network configurations of gravitational wave detectors, based on the projections in [29]. We start with the O1 sensitivity of the aLIGO Hanford and Livingston detectors as a comparison point [25], then add the Virgo detector with an initial noise curve as projected in [29]. We compare this configuration to the network of Hanford, Livingston and

Virgo (HLV) at design sensitivity [50, 53], and with a network expanded to include LIGO India [55], KAGRA [54] and both. Over the lifetime of the second generation instruments we expect the performance of the global network to improve parameter estimation in three important ways. The expansion of the global detector network should give better sky resolution and ability to better measure the signal polarisation (the Hanford and Livingston detectors are nearly co-aligned); the improvements at low frequency will increase the observable duration of the signals and lead to more cycles of the inspiral part of the waveform being observable; and finally the overall decrease in noise levels will greatly increase the signal-to-noise ratio of the source. We investigate the effect of these improvements on sky localisation, mass measurement, and distance and inclination accuracy for a GW150914-like system.

2.2 Method

To compare the different network configurations we use a set of sixteen simulated signals and perform the full parameter estimation for each network set-up and each signal. We used the reduced order model of the SEOBNRv2 waveform [67, 68], which includes aligned spins on both component bodies, for both simulation and recovery of the source parameters. The signal parameters were chosen to lie within the posterior distribution for the GW150914 event, so our simulations will appear to have the same relative amplitude in each detector as GW150914 did. This allowed us to easily verify that the results appeared similar to GW150914 when using the O1 (2015-2016) HL detector network.

We use a set of different network configurations which are designated by an identifier with three parts: The detector network, lower cut-off frequency, and noise spectrum. The detector network setup is a combination of aLIGO Hanford (H), aLIGO Livingston (L), AdVirgo (V), LIGO-India (I), and KAGRA (J). The noise spectrum is labeled as either “Early” , which indicates empirical ER8b/O1 (The first observing run of aLIGO and the eighth engineering run immediately preceding it, i.e. September 2015; PSD estimated from data taken shortly before the arrival of GW150914) spectra for H and L, and the projected early low curve for Virgo [29], or “Design” for the expected sensitivities at final design specification [29, 54]. The lower

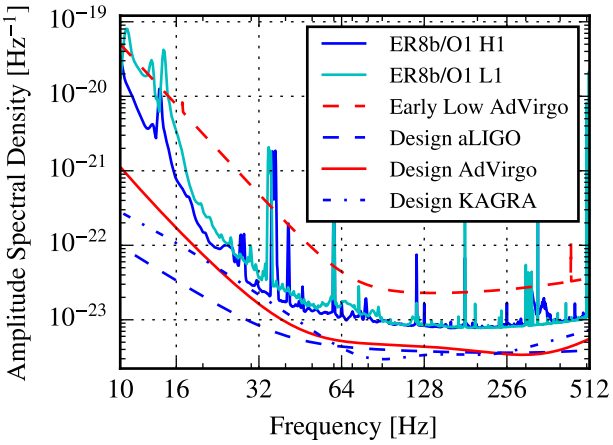

Figure 2.1: The amplitude spectral densities for the different sensitivities. The jagged curves for the ER8b/O1 curves are the median time averaged noise around GW150914 [69]. All other curves are projected sensitivities [29].

cut-off frequency is either 10 Hz or 30 Hz as little signal power is accumulated below 30Hz for the low sensitivity cases. For all sensitivities the noise levels below 10 Hz rise very sharply, as can be observed in Figure 2.1. We selected twelve combinations of these which are shown in Table 2.1.

We neglected 10 Hz runs for early networks as they would yield little benefits due to the high noise levels at low frequency. KAGRA and LIGO India are still in construction phase and cannot be expected to start observing for some years, therefore the HLV detectors form the basis of the runs at design sensitivity. As LIGO India is still under construction we assume the arms to be aligned to North and East. Two 2-detector runs are included to represent the minimal possible configuration and to give results comparable to GW150914 [51].

The amplitude spectral densities for all sensitivity curves are given in figure 2.1. The prior range of the component masses of the binary is $10-80\ M_{\odot}$ so that it is wide enough to contain all possible simulations drawn from the GW150914 posterior and the posterior of those systems. Similarly, we chose to set the segment length to 8 s for 30 Hz runs and 160 s for the 10 Hz runs, based on the time spent in the analysed frequency band combined with a safety margin. We chose a zero noise realization to avoid

noise perturbations affecting the comparisons. We assume perfect phase and amplitude calibration, although we consider the effect of an uncertainty on calibration of 10% in amplitude and 10° in phase in the appendix. All parameter estimation was performed using LALInference [41] in its nested sampling mode.

2.3 Results

Individual parameters are characterized by different features of the waveform, and therefore affected differently by the improvements in noise levels or network extensions. This reflects the distinction between intrinsic parameters, which are properties of the source itself, and extrinsic parameters, which are related to the relative positions and orientations of the source and the detectors. For this reason we present the results for different parameters in their individual sections, which also include the discussion of the results, and their comparison to the expected scaling relations. Table 2.1 contains an overview of the results for all discussed parameters with each run averaged over all simulations.

The quantities used to measure the precision of the parameter estimation are the sizes of the 90% credible interval or area for the chirp mass, distance, and sky area, and the value from the maximum likelihood sample for the signal-to-noise ratio.

2.3.1 Signal to Noise Ratio

We use the optimal signal-to-noise ratio (SNR) as a metric for comparing the strength of a signal against the background, which we define as $\sqrt{\langle h|h\rangle}$, where $\langle\cdot|\cdot\rangle$ is the noise-weighted inner product as defined in Equation 1.35. In practice the integral covers a finite range $[f_{\min}, f_{\max}]$, where $f_{\min}$ is the low frequency cut-off, which is chosen according to the noise properties so that the signal does not accumulate significant SNR below that value. $f_{\max}$ is chosen to be above the highest frequency contribution in the signal. This relation shows that both lowering the cut-off frequency $f_{\min}$ and decreasing the noise $S(f)$ improve the SNR by either increasing the interval over which the SNR can be accumulated, or increasing the integrand itself. The amount to which these increase the value depends on the noise spectrum in the region of interest. Observing a signal in N detectors is expected to increase the

	SNR at $\mathcal{L}_{\max}$	Area [deg^2] 90% C.A.	$\Delta_{\mathcal{M}}$ [$M_\odot$] 90% C.I.	Δ_D [Mpc] 90% C.I.	$\Delta_{\theta_{\rm JN}}$[rad] 90% C.I.
GW150914 (final) [52]	23.7	230	3.9	340	—
Early HL 30	26.7 ± 1.6	183 ± 34	5.0 ± 0.5	306 ± 30	—
Early HLV 30	27.6 ± 1.7	8.4 ± 5.06	4.9 ± 0.5	237 ± 48	—
Design HL 10	75.9 ± 4.8	31 ± 7	0.24 ± 0.03	215 ± 39	—
Design HL 30	72.5 ± 4.7	34 ± 8	2.1 ± 0.4	218 ± 38	—
Design HLV 10	86.9 ± 7.2	0.57 ± 0.20	0.23 ± 0.03	179 ± 42	0.85 ± 0.15
Design HLV 30	84.1 ± 7.5	0.54 ± 0.18	1.8 ± 0.4	179 ± 42	0.86 ± 0.15
Design HLVI 10	99.9 ± 15.6	0.21 ± 0.15	0.20 ± 0.03	140 ± 56	0.65 ± 0.19
Design HLVI 30	96.4 ± 15.3	0.19 ± 0.15	1.6 ± 0.4	140 ± 55	0.65 ± 0.20
Design HLVJ 10	108 ± 14	0.14 ± 0.10	0.21 ± 0.03	95 ± 53	0.46 ± 0.26
Design HLVJ 30	107 ± 14	0.13 ± 0.10	1.5 ± 0.3	98 ± 51	0.46 ± 0.25
Design HLVIJ 10	119 ± 20	0.11 ± 0.08	0.19 ± 0.03	90 ± 54	0.44 ± 0.25
Design HLVIJ 30	116 ± 20	0.10 ± 0.08	1.4 ± 0.3	92 ± 53	0.44 ± 0.26

Table 2.1: This table contains the means of the corresponding values over all simulations with the standard deviation across the 16 values. The 90% credible areas were computed using the Skyarea Python module [70]. The SNR values are taken from the maximum likelihood sample of each posterior distribution. For chirp mass, distance and inclination we give the sizes of the 90% credible intervals $\Delta_{\mathcal{M}}$, Δ_D and $\Delta_{\theta_{JN}}$ respectively. $\mathcal{M}$ is defined in the detector frame. We omit the inclination angle values for the early networks and the 2-detector configurations, since the bi-modal posteriors are not well described by the 90% credible interval (see fig 2.5). GW150914 is included for comparison but needs to be used with care as those results assume a waveform low-frequency cut-off of 20 Hz, and allow a calibration uncertainty of 4.8% and 8.2% in amplitude and 3.2° and 4.2° in phase for Hanford and Livingston respectively [52].

SNR by a factor of $\sqrt{N}$ relative to using a single detector only, and not taking sensitivity patterns into account, since SNR adds in quadrature for independent measurements.

Figure 2.1 shows that the noise levels start to rise quickly for frequencies lower than $\approx$50 Hz for all sensitivities. This explains why we see only minor differences in the SNR between 10 Hz and 30 Hz runs, which increases by a factor of only $1.01-1.05$. When increasing the detector sensitivity to the full design sensitivity however the SNR increases by a factor of $\approx 2.7-3.0$. The gains from adding Virgo to the network in the low sensitivity case are minor, with a factor of 1.03. This is, again, expected as the early low AdVirgo sensitivity is significantly less than that of the aLIGO detectors so it does not contribute much to the SNR. In the high sensitivity case the difference is noticeable with Virgo increasing the SNR by a factor of ≈ 1.2, bringing the total to ≈ 84 for the whole network. The fourth detector increases the combined SNR by a factor of ≈ 1.2 for LIGO India and ≈ 1.3 for KAGRA which suggests that KAGRA was in a more advantageous position for this event. Adding both LIGO India and KAGRA to the 3 detector setups brings the total SNR to ≈ 116, which is ≈ 1.4 times higher than the three detector value. The gains are roughly compatible with the expected values derived above, though we would not expect an exact match as the argument neglects differences in noise spectra and the impact of the antenna patterns for the different detectors. The measured SNRs for one simulation are shown in figure 2.2.

2.3.2 Sky localisation

The sky localisation is mainly determined by the timing measurements between the individual detectors [38, 29]. This means that there are two components to the measurement: the layout and synchronization of the detectors, and the measurement of the time delay using this external information. The layout of the detectors is important in that errors in the time measurements can be mitigated if the light travel time between the detectors is maximized, as this reduces the relative impact of a fixed timing error. The timing accuracy is inversely proportional to both SNR and the effective bandwidth [38]. For small areas we can approximate the relevant section of the sphere as being flat, therefore the localisation is proportional

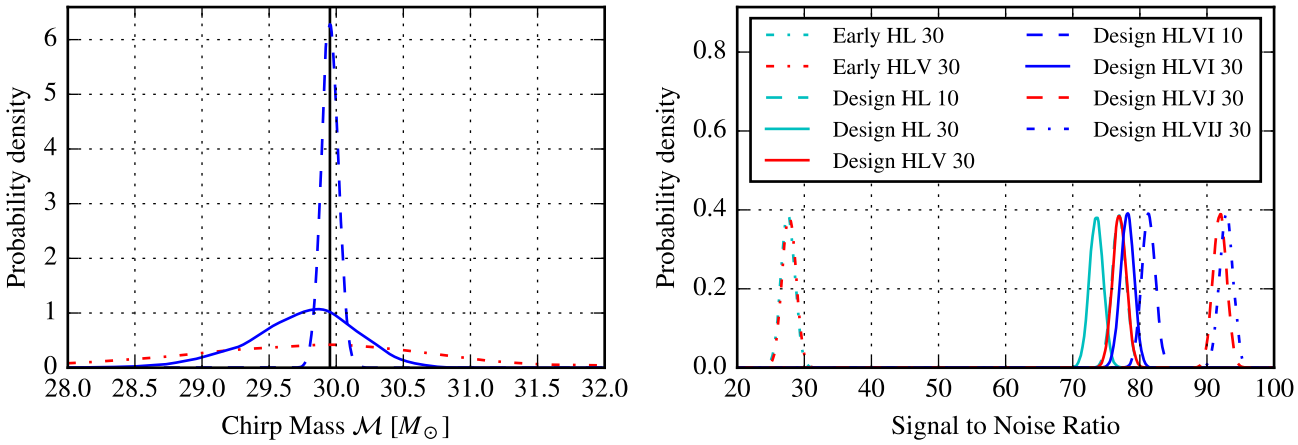

Figure 2.2: The posterior distribution of the chirp mass and signal-to-noise ratio for one individual signal. For the chirp mass only three runs are shown since the only factors which affect the distribution in a noticeable way are the switch to design sensitivity (red to solid blue), and the lowering of the lower cut-off frequency (solid blue to dashed blue). The main change in the signal-to-noise ratio is caused by the switch to design sensitivity detectors, although additional detectors and a lower cut-off do have a noticeable impact.

to the square of the timing error, so we get:

$$\sigma_{\text{area}} \propto \sigma_{\text{RA}}\sigma_{\text{Dec}} \propto \text{SNR}^{-2} \tag{2.1}$$

Even assuming perfect measurements, the nature of triangulation limits our ability to localize the source. Using only triangulation, with two detectors the source can be constrained to a circle, with three detectors to two points, and only the fourth detector allows us to narrow to location down to a single point. As the measurements are not perfect we do, however, still expect improvements from additional detectors beyond the fourth. Due to the fact that adding detectors does not only provide additional baselines for triangulation but also increases the SNR (see section 2.3.1), we expect massive improvements in the sky localisation when detectors are added to the network. These gains should be the highest for the third detectors as it reduces the annulus to two single points, and to a lesser degree from the fourth detector which breaks the last degeneracy stemming from the symmetry under reflections on the plane of three detectors. Another advantage of an expanded

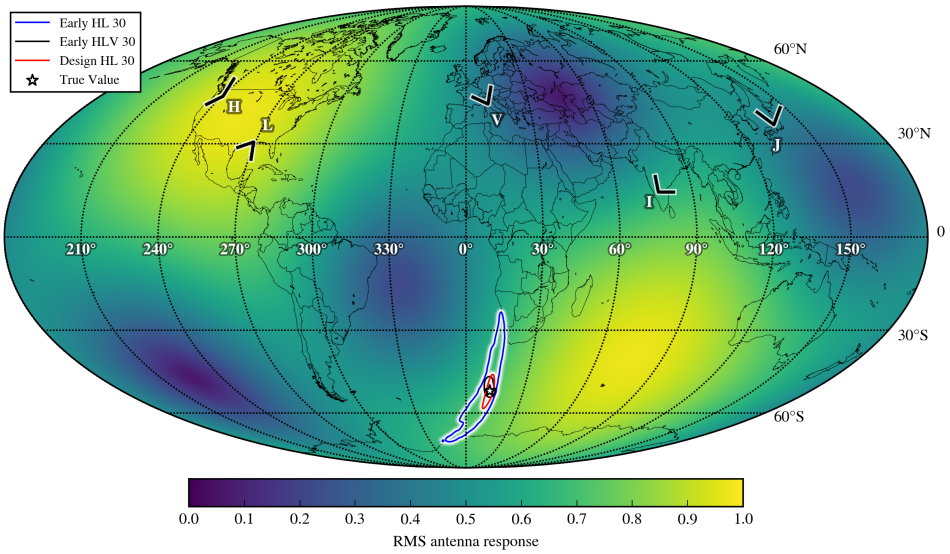

Figure 2.3: The 90% credible areas of one individual signal, over the average sensitivity patterns of the Hanford and Livingston detectors added in quadrature. The two yellow areas mark the locations with the highest antenna response while signals from sources located in the dark blue regions are strongly suppressed. The two large contours (blue and red) show the localisation with only two detectors, where the 10Hz design sensitivity run was skipped since it is indistinguishable from the 30Hz run. The smallest, black contour represents the ER8b/O1 configuration combined with the AdVirgo at its early low sensitivity. All 3+ detector runs at design sensitivity produce areas smaller than the marker for the true position. Figure 2.4 shows the central region magnified so those details become visible.

detector network is rooted in the non-uniform antenna pattern of gravitational wave detectors which is shown in figure 2.3. This causes detectors to have "blind spots" with low sensitivity, which can be compensated for by carefully choosing the position and orientation of other detectors. This helps to provide uniform sensitivity across the sky, and could increase the chances of making prompt electromagnetic follow-up observations of sources [71].

In addition to the timing triangulation, the relative amplitudes of the source in each detector, as determined by the angle-dependent antenna response functions, provides additional information about the position of the source which is naturally incorporated in our coherent analysis. This can break the ring-like or bimodal degeneracy in the two or three detector cases.

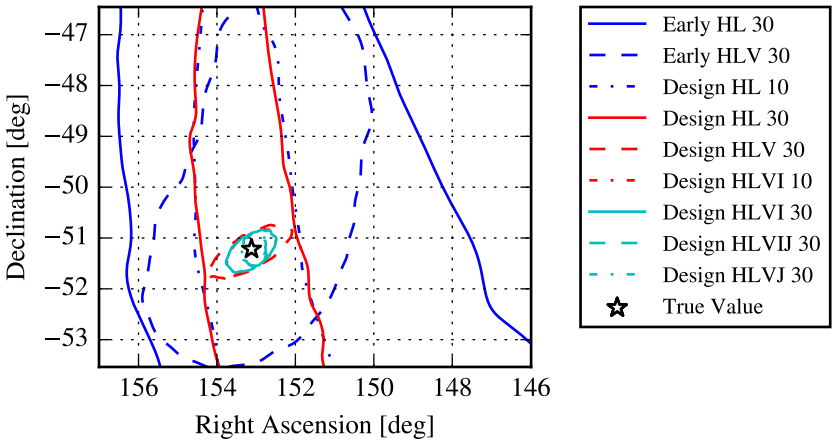

Figure 2.4: The magnification of the region surrounding the true position. While figure 2.3 shows the large scale localisation, this plot illustrates the differences between 3, 4, and 5-detector set-ups, which continuously shrink the area while remaining centred on the true location. Noteworthy is also that even at low sensitivity AdVirgo is able to shrink the area massively and collapse the annulus into a region with the diameter comparable to the width of the 2-detector ring. The increased sensitivity has a greater impact in the 3-detector set-up as the improvements in AdVirgo are much larger.

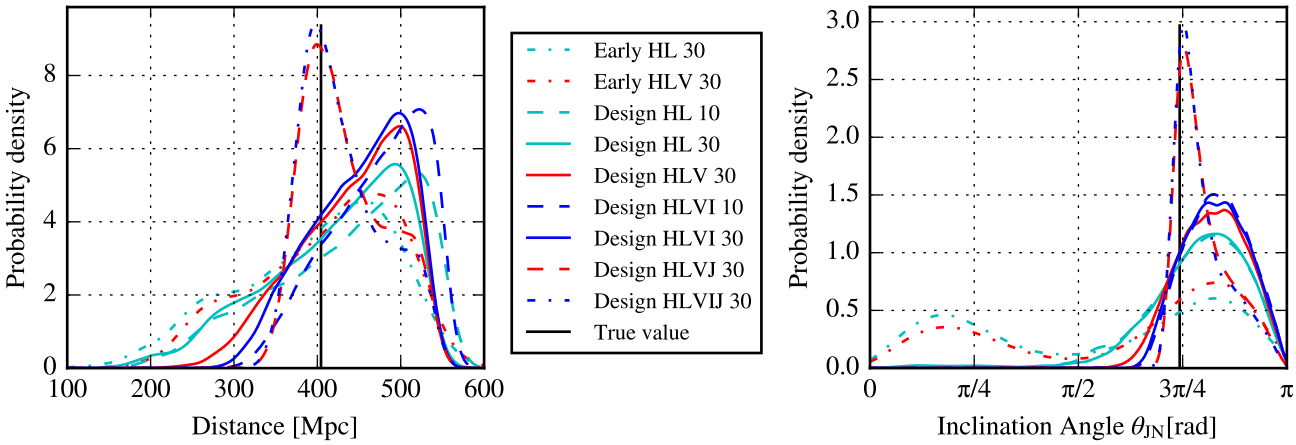

Figure 2.5: The posterior distributions of the luminosity distance to the source and the inclination angle, for one simulated signal. The distance measurement covers a wide range of values, as the distance prior is uniform in volume and the distance is not very well measured and degenerate with other extrinsic parameters. The inclination angle is only weakly constrained at current sensitivities or with only two detectors. While the degeneracy between face-on and -off orientations can be broken with a third detector at higher sensitivities, the width of the peak decreases only from $\approx \pi/4$ to $\approx \pi/7$. Qualitative differences with the posterior peaking close to the maximum appear only only once a fourth detector in an advantageous location is added.

We observe that adding AdVirgo to the two aLIGO detectors improves the localisation by factors of $\approx$22 and $\approx$64 for ER8b/O1 and design sensitivities respectively. Adding a fourth detector decreased the area by a factor of $\approx$2.8 when adding LIGO India or by $\approx$4.0 in the case of KAGRA over the HLV setup. The difference between these two possible 4 detector configurations is due to differences in sensitivity, as well as the antenna pattern. The full 5-detector configuration yields an area of ≈ 0.1 deg^2 on average, which is smaller than the 3-detector result by a factor of $\approx$5.3. The areas range from tenths to hundreds of square degrees and are given in Table 2.1.

Unexpectedly, for the 3+ detector networks lowering the cut-off frequency did not improve the measurement but worsened it, increasing the area by factors of $1.06 - 1.10$. This is due to the exclusion of low distances, which shifts the posterior into a region where the sky localisation is less precise. This is shown in figure 2.6 for the right ascension, though

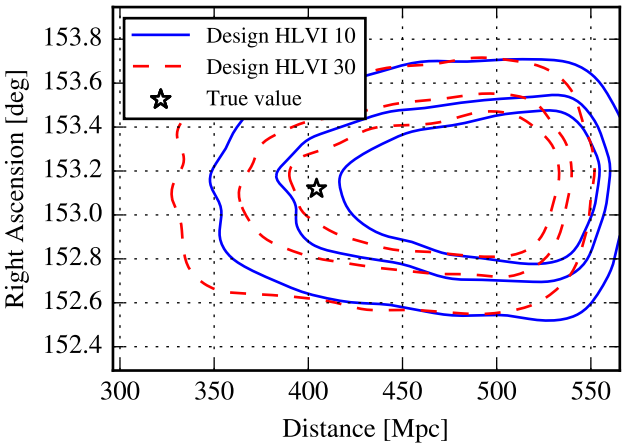

Figure 2.6: The combined posterior distribution of the right ascension and luminosity distance shows that constraining the distance to higher values increases the average width of RA posterior distribution. This behaviour is identical for the declination and thereby causes the localisation to worsen as low distance regions are excluded. The three different lines in either solid-blue or dashed-red show the 50%, 68%, and 90% percentiles.

the declination behaves identically. The distance shifts towards higher values instead of only narrowing as the distance prior is uniform in volume, therefore tighter constraints on the distance have a larger effect in the low distance region. In case of the 2-detector set-up we observe an improvement by a factor of 1.08.

The differences are illustrated by figure 2.3 for the large scale improvements, and figure 2.4 for the 3+ detector set-ups. Figure 2.3 additionally shows the sensitivity pattern of the combined H and L detectors at the time of detection.

2.3.3 Chirp Mass

The chirp mass, defined as $\mathcal{M} = (m_1 m_2)^{3/5}(m_1 + m_2)^{-1/5}$ is the most important quantity in determining the frequency evolution for a gravitational wave from compact binaries. Accordingly $\mathcal{M}$ can be measured precisely from the phase evolution of the waveform [26], in contrast to extrinsic parameters such as the distance, which are measured from the signal amplitude as measured in multiple detectors. Since the phase evolution is nearly identical in all de-

tectors, if each detector were equally sensitive the measurement error would be expected to scale as $\frac{1}{\sqrt{N}}$ with N being the number of detectors [26]. More generally, the measurement of the chirp mass improves due to the signal-to-noise ratio according to the following relation for post-Newtonian inspiral signals

$$\Delta(\ln \mathcal{M}) \propto \mathrm{SNR}^{-1}\mathcal{M}^{5/3} \tag{2.2}$$

[26]. We therefore expect the improved sensitivity to be helpful since it reduces the relative obfuscation of the waveform due to noise, increasing the signal-to-noise ratio. Additionally, the sensitivity improvement at low frequencies, allowing for a reduced lower frequency limit for the observed signal, is expected to be beneficial as it enables us to detect additional cycles of the inspiral which contain information about the chirp mass.

The results for the full set of networks considered are shown in Table 2.1. We report the detector-frame chirp mass measurements, which are affected by the red-shift of the source, but are the most easily comparable when looking at multiple systems which appear similar to the detectors. We find that with the ER8/O1 HL sensitivity the 90% credible interval in $\mathcal{M}$ was a mean of $5.0\,\mathrm{M}_\odot$, which is slightly higher than the range of $3.9\,\mathrm{M}_\odot$ reported in [52] for GW150914 using the SEOBNRv2 model, although this can be partly attributed to our use of $f_{\min} = 30\,\mathrm{Hz}$ as opposed to 20 Hz.

When adding detectors we see minor gains, improving the chirp mass estimate by factors of ≈ 1.02 and $\approx 1.04 - 1.15$ per detector added, for Early and Design sensitivity runs respectively. We attribute this to the relative sensitivity of the aLIGO and AdVirgo instruments, such that the SNR increases less than the $\sqrt{N}$ formula implies. Improving the sensitivity proves much more rewarding, yielding an improvement factor of $\approx 2.4 - 2.7$ when using the HL or HLV set-up at 30 Hz. The gains with a lowered frequency cut-off are even higher, improving the measurements by factor of $\approx 7.3 - 8.7$. Figure 2.2 shows a representative for each of the three distinct groups with nearly identical distributions. These groups are composed of the high cut-off, low sensitivity runs in the very wide case, the design sensitivity 30 Hz runs for the intermediate peak, and the sharply peaked results from the two 10 Hz runs.

2.3.4 Distance and Inclination

The inclination angle is the angle between the line of sight between the source and the observer $\vec{N}$, and the vector of the orbital angular momentum $\vec{L}$, which is aligned with the total angular momentum $\vec{J}$ in the aligned spins case considered here. It is a parameter which is typically weakly constrained by the gravitational wave observations, since it affects the relative amplitudes of the + and × polarisations which are not individually resolvable by a single interferometer. Restricting to the dominant $l = m = 2$ mode, the signal observed by the detector can be written as [72]

$$h(t) = \frac{1}{2}(1 + \cos^2(\theta_{JN}))F_+ A(t)\cos\Phi(t) + \cos(\theta_{JN})F_\times A(t)\sin\Phi(t), \quad (2.3)$$

with θ_{JN} being the inclination angle, $A(t)$, $\Phi(t)$ the amplitude and phase of the gravitational wave, and F_+, $F_\times$ the detector response functions for the + and × polarisations, which depend on the relative position and polarisation of the source (see Fig. 2.3). As the two aLIGO detectors are nearly co-aligned they cannot on their own resolve both polarisations very well, leading to a degeneracy between left and right elliptically polarised waves, i.e. under the transformation $\theta_{JN} \mapsto \pi - \theta_{JN}$. As the amplitude $A(t)$ is inversely proportional to the luminosity distance between source and observer, there is a further relationship between the inclination angle and the distance which allows edge-on nearby sources to appear similar to distance face-on (or face-off) sources. Together, these degeneracies produce the characteristic V-shaped posterior distributions as shown in e.g. Fig. 2 of [51], and the bimodal θ_{JN} marginal distributions shown in the right panel of Fig. 2.5 for the HL networks and for the Early HLV network.

While the inclination angle itself has little physical importance, the distance is important not only for the 3D source localisation, but also for the measurement of the masses in the source frame which needs to take the cosmological red-shift into account. This effect is already significant for GW150914 with a red-shift of only ≈ 0.1 and will only become more important for future detector networks, and especially third generation networks [73, 74] as higher sensitivities greatly increase the number of observable sources at high distances.

Figure 2.5 shows the posterior distribution for these two related parameters, with numeric values available in Table 2.1. The main feature is that

both parameters are only weakly constrained for all network configurations. We found that not only is the two-detector early HL network unable to break the degeneracy between θ_{JN} and $\pi - \theta_{JN}$, but the Early AdVirgo detector was not sensitive enough in comparison to Early aLIGO to do this either. At design sensitivity most but not all of the signals were isolated to one of the θ_{JN} modes. As soon as 3+ design sensitivity detectors are available the degeneracy was broken. The main difference aside from the degeneracy breaking we observe a qualitative change with the addition of KAGRA. This improves the width of the 90% credible interval by a factor of $\approx 1.5 - 1.8$ and shifts the peak to the true value. This behaviour is only observed without calibration uncertainty and discussed further in the appendix.

As there is an inverse relationship between distance and the signal amplitude, which is the quantity that is actually measured by the detectors, one might expect the uncertainty on distance to scale as $\Delta D/D \propto \mathrm{SNR}^{-1}$, by analogy to the Fisher matrix calculations for amplitude in [26]. However due to the correlations, the improved and extended detector network has a far lower effect on this set of parameters as compared to mass parameters. The size of the 90% credible intervals for distance and inclination respectively decrease from $\approx$306 Mpc and $\approx \pi/4$ for the ER8/O1 2-detector network to $\approx$90 Mpc and $\approx \pi/7$ for the complete design sensitivity network.

In a fashion similar to the slight worsening of the sky localisation, the size of the 90% credible distance interval does not always decrease when switching to a 10 Hz lower frequency cut-off. This is also caused by the small shift to higher distances, although the relative errors do decrease slightly as expected.

2.4 Effect of Calibration Uncertainty

In addition to the main network properties investigated above we replicated the analysis with a calibration uncertainty of 10% in amplitude, and 10° in phase, using the same interpolating spline model [75] as used in [51], which is a conservative estimate of the uncertainty that may be expected for on-line calibration (and therefore relevant for initial parameter estimates) [76]. At 10 Hz we used only the HL and HLVI configurations due to large parameter spaces and consequent resource consumption.

The most significant differences appear in the extrinsic parameters. The

sky localization worsens by a factor of $2.6 - 3.3$ for the early networks, and $7.5 - 10$ at design sensitivity. For distance and inclination angle the calibration improves the constraints by factors of $1.2 - 1.4$ for the HLVI configuration and 1.6 to 1.8 for networks including KAGRA. The observation of a sharp peak around the true value of both inclination and distance is a feature that starts to appear for 4+ detector design sensitivity networks. It is absent when using the 10%/10° calibration uncertainty. The chirp mass measurement is affected to a much smaller degree. It worsens by factors of 1.1 for early networks and $1.2 - 1.4$ at design sensitivity. The changes to signal-to-noise ratio are on a level below 1%.

	SNR at $\mathcal{L}_{\max}$	Area [deg^2] 90% C.A.	$\Delta_{\mathcal{M}}$ [$M_\odot$] 90% C.I.	Δ_D [Mpc] 90% C.I.	$\Delta_{\theta_{\rm JN}}$[rad] 90% C.I.
GW150914 (initial) [51]	25.1	610	4.4	350	—
Early HL 30	26.7 ± 1.8	599 ± 78	5.4 ± 0.4	327 ± 25	—
Early HLV 30	27.4 ± 1.7	22 ± 6	5.3 ± 0.4	256 ± 48	—
Design HL 10	75.7 ± 4.7	277 ± 33	0.33 ± 0.04	302 ± 25	—
Design HL 30	72.3 ± 4.6	262 ± 33	2.6 ± 0.5	295 ± 30	—
Design HLV 30	84.1 ± 7.4	5.5 ± 0.9	2.3 ± 0.4	219 ± 43	—
Design HLVI 10	100 ± 16	1.7 ± 0.5	0.26 ± 0.04	191 ± 38	0.87 ± 0.06
Design HLVI 30	96.3 ± 14.8	1.4 ± 0.6	2.0 ± 0.4	192 ± 35	0.88 ± 0.05
Design HLVJ 30	107 ± 14	1.2 ± 0.5	2.0 ± 0.4	167 ± 43	0.76 ± 0.18
Design HLVIJ 30	116 ± 20	0.85 ± 0.27	1.8 ± 0.4	163 ± 43	0.74 ± 0.19

Table 2.2: This table is structurally identical to Table 2.1 while presenting the results of the analysis using a calibration uncertainty of 10% in amplitude and 10° in phase. It contains the means of the corresponding values over all signals with the standard deviation across the 16 values. The 90% credible areas were computed using the `skyarea` Python module [70]. The SNR values are taken from the maximum likelihood sample of each posterior distribution. The values given for the chirp mass $\mathcal{M}$ and distance are the sizes of the 90% credible interval. $\mathcal{M}$ is defined in the detector frame. We omit the inclination angle values for the two and three detector configurations, since the bi-modal posteriors are not well described by the 90% credible interval (see fig 2.5). In the first row we include the first GW150914 results, reported in [51], which includes comparable calibration uncertainty of 10% in amplitude and 10° in phase for both Hanford and Livingston. Note that this is different from the GW150914 (final) in Table 2.1.

Numeric values for all runs including calibration uncertainty are given in Table 2.2. We observe that, while the improvement of individual detectors

and expansion of the network are important, improving the calibration is essential to obtaining the best possible results from the available detectors.

2.5 Conclusion

Although based around the GW150914 system, the results presented here give a good indication of the qualitative behaviour of parameter estimation for binary black hole systems as the global network of gravitational wave detectors continues to expand and improve in sensitivity. A less extensive subsequent analysis found that using the same procedure on GW151226 produces comparable results where expected. There are minor differences caused by the lower mass which gives more importance to the inspiral over the merger and ring-down, as noted in [52]. A large difference was observed in the ratios of sky areas caused by the initially poor localisation of the early HL network.

While the observed improvements in chirp mass were comparable to the predictions from Fisher matrix calculations, they tend to under-perform slightly. This is expected since the detectors are not identical and the noise curves differ, especially at the low frequency end which is relevant for the chirp mass measurement. For distance and inclination, the behaviour is poorer due to the correlation and degeneracy between parameters, and we confirm that the naive scaling relationship based on the Fisher matrix does not hold, even for signal-to-noise ratios of 26 and above. Instead, the greatest effect comes from the expansion of the network and elimination of a large region of the sky, and the relative geometry of source and detectors. In general we expect the breaking of degeneracies to play an important role, but one that can vary significantly between different sky positions, as the relative detector responses change the amplitude of the signal in each detector.

With the combined improvements in sky localisation and distance measurement the volume to which future coalescence events will be constrained can be expected to decrease substantially as detectors are added and improved. As soon as a third detector joins the network the area which needs to be covered by EM observers decreases by factors of 20-60 which will allow for a more complete coverage and greater depth to increase the chance of observing potential counterparts, or the (statistical) identification of a host galaxy [77, 78, 79]. In fact, there has been a three-detector BBH detec-

tion since this work has originally been completed, GW170814 [39], which yielded a 90% confidence area of $\approx 60\text{deg}^2$ at a distance of 540^{+130}_{-210}Mpc and a SNR of 18, which is slightly less well localised than we expected based on our simulations in the chapter. Such differences are to be expected as source properties and sky location differ between the two events. Breaking the distance-inclination degeneracy will also aid the ability to perform cosmology with gravitational wave sources [80, 81, 82, 83, 84]. A comparison between the precise and imprecise calibration results in Tables 2.1 and 2.2 also show the importance of precise calibration in estimation of the distance, even with a global network of four or more detectors. For more detailed knowledge about the intrinsic properties of the sources themselves, the main driver is the improvement of the sensitivity. In case of the chirp mass the most important region is at low frequency where a factor of 3 in the cut-off can tighten the constraints by an order of magnitude.

In summary, although the field of gravitational wave astronomy as a true observational science has only just begun, the currently planned upgrades and expansions of the global network of detectors offer good observational prospects for heavy stellar mass binary black holes such as GW150914. Our work highlights the differing roles of (low-frequency) sensitivity and network geometry in aspects of constraining the source properties, indicating that a global network of comparable detectors will be necessary to achieve the best results for both mass estimates and source localisation.

Chapter 3

Digging the population of compact binary mergers out of the noise.

Adapted from [5] with minor changes. My contribution to this work was to a) *derive the model,* b) *run the simulations,* c) *produce most of the plots and text. This paper was published in* Mon. Notices Royal Astron. Soc. *in April 2019.*

3.1 Introduction

Since the first detection of gravitational waves in 2015 [1], the Advanced LIGO and Advanced Virgo detectors have observed the coalescence of multiple compact binary systems, and have begun to reveal the population of coalescing compact objects [85]. This population is enabling studies in fields from probing alternative theories of gravity to constraining models of stellar evolution. These tend to be interested either in individual, preferably loud, signals, or in the population of sources as a whole. The latter type of population analysis tries to estimate the parameters governing the distribution of sources in the Universe; their masses and spins, and the value of the astrophysical merger rate [56] are of particular interest. As the sensitivity of detectors improves over the coming years, the detected number of sources is expected to grow at an accelerated pace, rapidly increasing the amount of information available for population studies [29].

When undertaking population analyses, one has to consider that real detectors may produce noise transients which cannot in all cases be distinguished from astrophysical GW sources. To avoid population inferences being biased by noise events one might consider only events with a much higher probability to be of astrophysical origin than to be caused by noise artefacts. In templated searches for compact binaries, the relative probability of astrophysical vs. noise origin for a candidate event is a function of a detection statistic calculated for each event by a search analysis pipeline (see e.g. Abbott et al. [18], Usman et al. [86], Cannon et al. [87], Cannon et al. [88], Messick et al. [89], Nitz et al. [90]). Typically a candidate event is generated as a local maximum in matched filter signal-to-noise ratio (SNR) above a search threshold; the detection statistic value then incorporates the matched filter SNR, as well as other goodness-of-fit tests to reject non-Gaussian instrumental noise transients [91, 92].

At low SNR, the population of events is dominated by the noise 'background', whereas at high SNR (or in general for events assigned high statistic values) the astrophysical 'foreground' dominates.[1] To limit possible pollution of the sample used for population inference, one may place a minimum threshold on the detection statistic; any event above threshold is then assumed to be astrophysical, whereas all other events are discarded as potential noise transients. Note that the choice of threshold value requires an empirical estimate of the rate and distribution of background events [93], since the rate, strength and morphologies of detector noise artefacts are not known *a priori* [94].

A simple strategy of thresholding is sub-optimal for two reasons. First, discarding events below the threshold will almost certainly discard information from some number of quiet but still identifiable signals [95]; second, there is still a finite chance that the resulting 'signal' set is nevertheless contaminated by noise, leading to potentially biased inferences. The choice of SNR threshold requires a trade-off between these two considerations and depends on the intended use. One also has to take into consideration any bias in the observed population produced by the effect of source parameters on the loudness of a signal, and thus its chance of exceeding a SNR threshold [96]; we expect potentially major observation selection effects for binary

[1] We will loosely refer to the detection statistic as 'SNR' when discussing the distinction between instrumental noise events and astrophysical events.

mass(es) [97] and component spins [98].

Here, we propose a method which alleviates the issues associated with simple thresholding by applying a hierarchical mixture model under which each event is considered to originate from either a foreground (astrophysical) or a background (noise) population. For each event, the probability of either case naturally defines a weight for its contribution to inferences on population parameters.

This method combines the processes of estimating the expected number of events in either class [99] (the number of foreground events being a proxy for the astrophysical merger rate density) and estimating parameters of the underlying populations, which have previously been performed separately. It avoids being biased through the inclusion of background events, while being able to make use of events with a non-negligible probability of noise origin, which would be discarded by thresholding. In theory this method allows the SNR threshold to be reduced to an arbitrarily low value, though in practice we are still limited by the computational resources required to extract the source parameters from each event under consideration.[2]

Our method is applicable to any hierarchical model of a source population, examples of which have been explored in the literature. This includes analyses which combine information from multiple events to infer a parameter common to all, such as deviations from general relativity [101] or a parameterised neutron star equation of state [102]. The use of a mixture model with astrophysical and noise populations is particularly useful when the model of interest has a strong effect on the detectability of sources, i.e. the detected events are unrepresentative of the underlying population. A good example is the mass distribution of sources, which we consider below.

For a compact binary in its inspiral phase, the frequency-domain amplitude of the gravitational wave in the stationary phase approximation is proportional to $\mathcal{M}^{5/6}$, where $\mathcal{M} = (m_1 m_2)^{3/5}(m_1 + m_2)^{-1/5}$ is the chirp mass of the signal [26]. Since the gravitational-wave detectors are sensitive to the signal amplitude, more massive sources will produce a larger SNR for a fixed position relative to the detectors. For a search which counts signals above a particular threshold, more massive signals will be over-represented in

[2] An analysis that effectively removes all SNR thresholds, applying Bayesian analysis to the entirety of the gravitational-wave data set rather than restricting to data close to events triggered on SNR maxima, is proposed in [100]; its application appears at present to be still more limited by computational cost.

the selected events by a factor $\approx \mathcal{M}^{5/2}$, neglecting cosmological red-shifting of the source and assuming a constant source rate per unit volume at all distances.

Messenger and Veitch [103] considered the problem of selection effects in mass distribution inference by dividing the observing time into discrete chunks, which each contain zero or one sources, and computing population likelihoods while accounting for false alarms and false dismissals from an idealised noise distribution. Farr et al. [99] derived an equivalent formalism for rate inference that allows a population shape function to be estimated alongside. Our derivation in Section 3.2 follows similar lines.

The selection function for masses was important in estimating the astrophysical event rates in the first Advanced LIGO observing run (O1), which inferred rates using a mixture model, for fixed choices of population shape (i.e. mass distribution) [56, 104]. A separate analysis also estimated the slope of a power law model of the mass distribution function considering the detected events, described in [85] (and updated in [39]).

The selection function in the form of a sensitivity-weighted measure of the space-time volume VT surveyed for signals above a certain SNR threshold is also important when considering searches which do not make a clear detection. There, the loudest background event, or a nominal detection threshold, is used to set an upper limit on the astrophysical rate of a fiducial source population, for example limits on the rate of mergers of binary neutron stars and neutron star–black hole binaries in O1 [105]. The sensitive-volume approach has been in use since the initial detector era [106, 107], and continues to be refined to incorporate mass- and spin-dependent selection effects [108, 109, 98] and cosmological effects, as well as to improve the accuracy of measurement [97].

As the number of detections increases, determination of the population of coalescing compact binaries is expected to provide insight into the astrophysics of black hole and neutron star binary formation [110, 111, 112]. Population synthesis models can describe the masses and spins of coalescing compact binaries under a variety of formation scenarios (see e.g. Belczynski et al. [113], Belczynski et al. [114, 115], Spera et al. [116]). Comparison of these predictions to the observed distribution can be used to constrain the uncertainties in parametrised models of source populations [117, 118]. This has motivated the development of methods to determine the mass-

dependent coalescence rate in the absence of false alarms, using both specific parameterised models [119, 120, 121, 122] and non-parametric methods [123]. The alignment of black hole spins is expected to be a key distinguishing feature between binaries formed in the field or through dynamical interactions (see e.g. O'Shaughnessy et al. [124], Mandel and O'Shaughnessy [125], Gerosa et al. [126], Stevenson et al. [127], Farr et al. [128], Stevenson et al. [129], Gerosa and Berti [130], Fishbach et al. [131], Talbot and Thrane [132], Vitale et al. [133]), which also requires an understanding of the spin selection function [108, 109, 98, 134]. This is caused by the 'orbital hang-up' effect [135], where binaries with component spins aligned with the total orbital angular momentum tend to inspiral (i.e. reduce orbital radius) more slowly than those with anti-aligned spins. This leads to an increase in the radiation emitted at specific frequencies in the sensitive band of ground-based detectors, thus increasing the detectability of these sources.

The work presented here is complementary to these studies, as it aims to incorporate an astrophysical distribution model as part of a mixture with a noise component. As the observed population is limited by the sensitivity of Advanced ground-based detectors, the population of candidate sources at the greatest distances (lowest detection significance) will be contaminated with background events. We expect our method, using information from such sources, to improve both the precision and accuracy of merger rate and population parameter estimates; though as we will see, the degree of improvement depends on how easily the foreground and background populations can be separated by existing analyses.

We start by defining our notation and deriving the general form of the model in section 3.2. Section 3.3 describes its application to a toy model of mass distribution inference in the presence of noise, and shows its application to a range of simple analytic population models. In section 3.5 we consider a more realistic simulated data set derived from an engineering run prior to the start of Advanced LIGO observations in 2015. We conclude in section 3.6.

3.2 Derivation of the generic model

We consider a mixture of two populations, the astrophysical 'foreground' and terrestrial noise 'background': quantities defined analogously for both populations will be distinguished by the subscripts F or B respectively. Quantities

without subscript then refer to the total population which is the union of foreground and background.

The model is also hierarchical: each event, if assumed astrophysical, has a set of intrinsic properties such as component masses and spins, which we collectively denote $\vec{\gamma}$. The distribution of these properties over each population is assumed to have a form described by a set of hyper-parameters. We do not have access to the 'true' values of properties for each event, only to a set of samples from a (typically Bayesian) estimate based on data around the event. These samples are derived under the assumption that the event is astrophysical, thus events that are in fact background will also be assigned parameter estimates.[3]

We then define the core quantities used in the following derivation as

- ρ_i, $\{\rho\}$: ranking statistic for one, resp. for all events in a given data set.
- $N_{\rm obs}$, $N_{\rm F,obs}$, $N_{\rm B,obs}$: observed number of events above a threshold $\rho_i > \rho_{\rm thr}$
- $N_{\rm exp}$, $N_{\rm F,exp}$, $N_{\rm B,exp}$: expected number of events with $\rho_i > \rho_{\rm thr}$, when modelling these as a Poisson process
- $\theta_{\rm F}$, $\theta_{\rm B}$: hyper-parameters which describe the shape of the foreground and background populations
- η_i, $\{\eta\}$: indicator variable showing whether any given event, resp. all events, belong(s) to the astrophysical ($\eta = F$) or to the noise population ($\eta = B$)
- $\vec{\gamma}_i$, $\{\vec{\gamma}\}$: vector of samples representing the parameter estimates (masses, spins, etc.) of one, resp. all events, under the assumption that events are astrophysical.

We wish to infer the joint posterior probability distribution of rates and population parameters for the two populations, given some events for which $\{\rho\}$ and $\{\vec{\gamma}\}$ have been determined by the search and parameter estimation stages of data analysis:

$$p(N_{\rm F,exp}, N_{\rm B,exp}, \theta_{\rm F}, \theta_{\rm B} | \{\rho\}, \{\vec{\gamma}\}, N_{\rm obs}). \tag{3.1}$$

[3] We do not, of course, know with certainty that any given event is background.

Using Bayes' theorem we can express the posterior distribution (3.1) in terms of prior and likelihood functions,

$$\begin{aligned} &p(N_{\mathrm{F,exp}}, N_{\mathrm{B,exp}}, \theta_{\mathrm{F}}, \theta_{\mathrm{B}} | \{\rho\}, \{\vec{\gamma}\}, N_{\mathrm{obs}}) \\ &\quad = \frac{p(\{\rho\}, \{\vec{\gamma}\}, N_{\mathrm{obs}} | N_{\mathrm{F,exp}}, N_{\mathrm{B,exp}}, \theta_{\mathrm{F}}, \theta_{\mathrm{B}}) p(N_{\mathrm{F,exp}}, N_{\mathrm{B,exp}}, \theta_{\mathrm{F}}, \theta_{\mathrm{B}})}{p(\{\rho\}, \{\vec{\gamma}\}, N_{\mathrm{obs}})} \end{aligned} \tag{3.2}$$

We drop the normalisation constant $p(\{\rho\}, \{\vec{\gamma}\}, N_{\mathrm{obs}})$ and factor out the likelihood for N_{obs} as being independent of the population hyper-parameters θ_{F} and θ_{B},

$$\begin{aligned} &p(\{\rho\}, \{\vec{\gamma}\}, N_{\mathrm{obs}} | N_{\mathrm{F,exp}}, N_{\mathrm{B,exp}}, \theta_{\mathrm{F}}, \theta_{\mathrm{B}}) \\ &\quad = p(N_{\mathrm{obs}} | N_{\mathrm{F,exp}}, N_{\mathrm{B,exp}}) p(\{\rho\}, \{\vec{\gamma}\} | N_{\mathrm{F,exp}}, N_{\mathrm{B,exp}}, \theta_{\mathrm{F}}, \theta_{\mathrm{B}}) \\ &\quad = \frac{N_{\mathrm{exp}}^{N_{\mathrm{obs}}} e^{-N_{\mathrm{exp}}}}{N_{\mathrm{obs}}!} p(\{\rho\}, \{\vec{\gamma}\} | N_{\mathrm{F,exp}}, N_{\mathrm{B,exp}}, \theta_{\mathrm{F}}, \theta_{\mathrm{B}}), \end{aligned} \tag{3.3}$$

where we use a Poisson likelihood for N_{obs} with a total expected number of events $N_{\mathrm{exp}} = N_{\mathrm{F,exp}} + N_{\mathrm{B,exp}}$. The second term, $p(\{\rho\}, \{\vec{\gamma}\} | N_{\mathrm{F,exp}}, N_{\mathrm{B,exp}}, \theta_{\mathrm{F}}, \theta_{\mathrm{B}})$, is the likelihood for the observed SNRs and parameter estimates, for the mixture model. We assume each event is conditionally independent given the population parameters, and so the joint likelihood is just the product of the likelihood for each one,

$$\begin{aligned} &p(\{\rho\}, \{\vec{\gamma}\} | N_{\mathrm{F,exp}}, N_{\mathrm{B,exp}}, \theta_{\mathrm{F}}, \theta_{\mathrm{B}}) \\ &\quad = \prod_i p(\rho_i, \vec{\gamma}_i | N_{\mathrm{F,exp}}, N_{\mathrm{B,exp}}, \theta_{\mathrm{F}}, \theta_{\mathrm{B}}). \end{aligned} \tag{3.4}$$

Now, we can split each of these into terms for the astrophysical and noise submodels by introducing an indicator variable $\eta_i \in \{F, B\}$, whose probability will depend on the rate parameters $N_{\mathrm{F,exp}}$ and $N_{\mathrm{B,exp}}$,

$$\begin{aligned} &p(\rho_i, \vec{\gamma}_i | N_{\mathrm{F,exp}}, N_{\mathrm{B,exp}}, \theta_{\mathrm{F}}, \theta_{\mathrm{B}}) \\ &\quad = p(\rho_i, \vec{\gamma}_i | \theta_{\mathrm{F}}, \eta_i = \mathrm{F}) p(\eta_i = \mathrm{F} | N_{\mathrm{F,exp}}, N_{\mathrm{B,exp}}) \\ &\qquad + p(\rho_i, \vec{\gamma}_i | \theta_{\mathrm{B}}, \eta_i = \mathrm{B}) p(\eta_i = \mathrm{B} | N_{\mathrm{F,exp}}, N_{\mathrm{B,exp}}) \\ &\quad = p(\rho_i, \vec{\gamma}_i | \theta_{\mathrm{F}}, \eta_i = \mathrm{F}) \frac{N_{\mathrm{F,exp}}}{N_{\mathrm{exp}}} + p(\rho_i, \vec{\gamma}_i | \theta_{\mathrm{B}}, \eta_i = \mathrm{B}) \frac{N_{\mathrm{B,exp}}}{N_{\mathrm{exp}}}, \end{aligned} \tag{3.5}$$

where the probability of each class is just the expected fraction of the total

number. Since this is a sum of probability densities, special care must be taken to ensure all terms are properly normalised, such that

$$\int_{\rho_{\rm thr}}^{\infty} \mathrm{d}\rho \int \mathrm{d}\vec{\gamma}\, p(\rho, \vec{\gamma}|\theta_\eta, \eta_i) = 1, \tag{3.6}$$

for $\eta = \mathrm{F}$ and $\eta = \mathrm{B}$, where $\rho_{\rm thr}$ is a minimum SNR value for which events are considered, either as a result of the event generation method or as a choice to limit computational costs. Neglecting this normalization would introduce an artificial preference for one component over the other. An extension to further sub-populations is simply achieved by including additional classes with their own rate and hyper-parameters.

Recombining the pieces, we can write the desired posterior in Eq. (3.1) as

$$\begin{aligned} &p(N_{\mathrm{F,exp}}, N_{\mathrm{B,exp}}, \theta_{\mathrm{F}}, \theta_{\mathrm{B}}|\{\rho\}, \{\vec{\gamma}\}, N_{\rm obs}) \\ &\quad \propto p(N_{\mathrm{F,exp}}, N_{\mathrm{B,exp}}, \theta_{\mathrm{F}}, \theta_{\mathrm{B}}) e^{-N_{\rm exp}} \\ &\qquad \times \prod_i \left[\, p(\rho_i, \vec{\gamma}_i|\theta_{\mathrm{F}}, \eta_i = \mathrm{F}) N_{\mathrm{F,exp}} + p(\rho_i, \vec{\gamma}_i|\theta_{\mathrm{B}}, \eta_i = \mathrm{B}) N_{\mathrm{B,exp}} \,\right]. \end{aligned} \tag{3.7}$$

This expression is similar to Eq.(21) from Farr et al. [99] with an explicitly added dependence on source parameter estimates. This implies that our formalism reduces to the Farr et al. [99] result as used by the LIGO-Virgo Collaboration to estimate binary black hole merger rates [85, 56], if the event distribution over mass or similar parameters is not free to vary.

The dependence on event parameters arises through the use of samples $\vec{\gamma}_i$, $i = 1 \ldots n$, drawn from the likelihood function of the data d for a given point in parameter space $p(d|\vec{\gamma})$. These allow us to evaluate the population likelihood function via marginalisation over the unknown true parameters, using the n samples to perform a Monte Carlo integral as in [136],

$$\begin{aligned} p(d|\theta_{\mathrm{F}}) &= \int p(d|\vec{\gamma}) p(\vec{\gamma}|\theta_{\mathrm{F}}) \mathrm{d}\vec{\gamma} \\ &= \langle p(\vec{\gamma}_i|\theta_{\mathrm{F}}) \rangle_{p(d|\vec{\gamma}_i)} \\ &\approx n^{-1} \sum_j p(\vec{\gamma}_j|\theta_{\mathrm{F}}). \end{aligned} \tag{3.8}$$

Samples from the likelihood therefore serve as a useful intermediate representation of the raw interferometer data d.

To obtain a quantity directly relevant for an astrophysical interpretation, the expected number can be transformed into the local merger rate R using the observing time T and the sensitive volume $V(\gamma)$:

$$R = \frac{N_{\mathrm{F,exp}}}{T \int \mathrm{d}\vec{\gamma}\, V(\vec{\gamma}) p(\vec{\gamma}|\theta_{\mathrm{F}})}, \tag{3.9}$$

where the integral marginalises over the space of source parameters $\vec{\gamma}$. In practice $V(\vec{\gamma})$ is estimated for a particular dataset by a Monte Carlo campaign, adding ('injecting') a large number of simulated signals to the data and counting the resulting number of events above threshold.[4]

An additional quantity which is not directly used in the derivation of our model, but is important for an astrophysical interpretation, is the probability of any given event originating from the astrophysical foreground, p_{astro}:

$$\begin{aligned} & p(\eta_i = \mathrm{F}, N_{\mathrm{F,exp}}, N_{\mathrm{B,exp}}, \theta_{\mathrm{F}}, \theta_{\mathrm{B}} | \rho_i, \vec{\gamma}_i) \\ = & \frac{p(\rho_i, \vec{\gamma}_i | \theta_{\mathrm{F}}, \eta_i = \mathrm{F}) N_{\mathrm{F,exp}} \times p(N_{\mathrm{F,exp}}, N_{\mathrm{B,exp}}, \theta_{\mathrm{F}}, \theta_{\mathrm{B}})}{p(\rho_i, \vec{\gamma}_i | \theta_{\mathrm{F}}, \eta_i = \mathrm{F}) N_{\mathrm{F,exp}} + p(\rho_i, \vec{\gamma}_i | \theta_{\mathrm{B}}, \eta_i = \mathrm{B}) N_{\mathrm{B,exp}}}, \end{aligned} \tag{3.10}$$

$$\begin{aligned} p_{\mathrm{astro},i} = & \iiiint p(\eta_i = \mathrm{F}, N_{\mathrm{F,exp}}, N_{\mathrm{B,exp}}, \theta_{\mathrm{F}}, \theta_{\mathrm{B}} | \rho_i, \vec{\gamma}_i) \\ & \mathrm{d}N_{\mathrm{F,exp}}\, \mathrm{d}N_{\mathrm{B,exp}}\, \mathrm{d}\theta_{\mathrm{F}}\, \mathrm{d}\theta_{\mathrm{B}}\,, \end{aligned} \tag{3.11}$$

where we marginalised over the population parameters. The integration range for both $N_{\mathrm{F,exp}}$ and $N_{\mathrm{B,exp}}$ is $(0, \infty)$, while the population parameters θ_{F} and θ_{B} are integrated over their respective domains.

3.3 Toy model

We construct a simple toy model of the Universe to test our inference framework in various ways. The toy model allows us to generate a large number of realisations from the same underlying parameters, and to be certain that we use the correct model when analysing these realisations. For simplicity we consider a static, flat, and finite universe. Events are characterised completely by the distance r to the source and a single mass parameter m,

[4] If the dataset already contains a number of detectable astrophysical signals, then the expected number of such GW events above threshold should be subtracted from the simulated event count. Alternatively, a data set which is, to a good approximation, empty of astrophysical signals may be used for the sensitivity estimate.

which takes the place of the γ used in the previously derived expressions. This mass parameter can be thought of as similar to the chirp mass. Additional effects such as inclination, spins, mass ratio, or antenna patterns are ignored.

For our detection statistic ρ we simply use (a simplified proxy for) the SNR, whose expected value ρ_{true} is determined by r and m as

$$\rho_{\mathrm{true}} = K\frac{m}{r}. \tag{3.12}$$

where K is an arbitrary constant which quantifies the detector sensitivity. We also model the uncertainty in the estimation of the mass parameter as

$$\sigma_{\mathrm{PE}} \propto \frac{m}{\rho} \tag{3.13}$$

which is a simplification of the relation given in [26].

To apply the generic form found in Eq. (3.7) to a specific problem, we need to evaluate the terms $p(\rho_i, \vec{m}_i|\theta_{\mathrm{F}}, \eta_i{=}\mathrm{F})$ and $p(\rho_i, \vec{m}_i|\theta_{\mathrm{B}}, \eta_i{=}\mathrm{B})$. This involves finding a functional form for the selection effects. For sources distributed uniformly in our static universe we can derive the needed expression directly by manipulating the joint distribution of masses and observed (detection) SNRs ρ_{obs}. We use the SNR relation defined above in Eq. (3.12), which defines a mass dependent lower cut-off $\rho_{\mathrm{cutoff}}(m) = K\ m\ r_U^{-1}$ to the ρ_{true} possible in our toy universe with radius r_U. Additionally we use the fact that the Euclidean distances r of sources distributed uniformly in volume follow a r^2 distribution. Therefore:

$$\begin{aligned}
p(\rho_{\mathrm{obs}}, m) &= \frac{1}{V_U}\int_0^{r_U} \mathrm{d}r\ 4\pi r^2 p(\rho_{\mathrm{obs}}, m|r) \\
&= \frac{1}{V_U}\int_0^{r_U} \mathrm{d}r \int_{\rho_{\mathrm{cutoff}}(m)}^{\infty} \mathrm{d}\rho_{\mathrm{true}}\ 4\pi r^2 p(\rho_{\mathrm{obs}}|\rho_{\mathrm{true}}) \\
&\qquad\qquad \times p(\rho_{\mathrm{true}}|m, r)p(m) \\
&= \frac{4\pi p(m)}{V_U}\int_{\rho_{\mathrm{cutoff}}(m)}^{\infty} \mathrm{d}\hat{\rho} \int_{\rho_{\mathrm{cutoff}}(m)}^{\infty} \mathrm{d}\rho_{\mathrm{true}}\ p(\rho_{\mathrm{obs}}|\rho_{\mathrm{true}}) \\
&\qquad\qquad \times \frac{(Km)^3}{\hat{\rho}^4}\delta(\hat{\rho} - \rho_{\mathrm{true}})
\end{aligned}$$

$$= \frac{4\pi K^3 m^3 p(m)}{V_U} \int_{\rho_{\rm cutoff}(m)}^{\infty} {\rm d}\rho_{\rm true}\, \rho_{\rm true}^{-4}\, p(\rho_{\rm obs}|\rho_{\rm true})$$
$$\propto m^3 p(m) \int_{\rho_{\rm cutoff}(m)}^{\infty} {\rm d}\rho_{\rm true}\, \rho_{\rm true}^{-4}\, p(\rho_{\rm obs}|\rho_{\rm true}), \tag{3.14}$$

where V_U is the volume of our universe, and $\rho_{\rm cutoff}(m) = \frac{Km}{r_U}$ is the lower SNR cut-off defined by a source of mass m being placed at the maximum allowed distance r_U. Thus, the SNR distribution for astrophysical sources is $p(\rho) \propto \rho^{-4}$, and we expect the observed mass distribution to be biased by a factor of m^3. The mass and SNR components of Eq. (3.14) are generally connected via the mass dependent SNR cut-off. The term $p(\rho_{\rm obs}|\rho_{\rm true})$ accounts for the shift in search SNR relative to the expected value due to detector noise:

$$p(\rho_{\rm obs}|\rho_{\rm true}) = \chi_{\rm NC}(\rho_{\rm obs}; \lambda = \rho_{\rm true}, k = 2), \tag{3.15}$$

where $\chi_{\rm NC}$ is the non-central chi distribution with a non-centrality of $\lambda = \rho_{\rm true}$ and $k = 2$ degrees of freedom.

For real binary merger events we can pursue an analogous derivation, though the resulting relation differs as the SNR is a more complex function of event parameters than Eq. (3.12). Additional complications arise if the detector is sensitive to events at cosmological distances, causing the observed masses to be redshifted by a distance-dependent amount.

The search and parameter estimation analyses that produce our events can only cover a finite range of masses, of which we denote the limits as $m_{\rm min}$, $m_{\rm max}$. We will assume that all astrophysical foreground events have masses lying within these limits; in practice one should take sufficiently wide limits that the density of foreground events at these limits becomes vanishingly small.

For the mass distribution of the astrophysical foreground we consider two types of population distribution: a truncated power law

$$p(m|\theta_{\rm F}) \equiv p(m|\alpha, m_{\rm low}, m_{\rm high}) \propto \begin{cases} m^\alpha & \text{if } m_{\rm low} < m < m_{\rm high} \\ 0 & \text{else} \end{cases} \tag{3.16}$$

with three free parameters, the slope α, lower mass cut-off $m_{\rm low}$, and high mass cut-off $m_{\rm high}$. The two mass cut-offs are constrained by the mass range

considered as $m_{\rm min} \leq m_{\rm low} < m_{\rm high} \leq m_{\rm max}$. The second population is a Gaussian

$$p(m|\theta_{\rm F}) = p(m|\mu,\sigma) \propto \begin{cases} \mathcal{N}(m;\mu,\sigma) & \text{if } m_{\rm min} < m < m_{\rm max} \\ 0 & \text{else} \end{cases} \tag{3.17}$$

with two free parameters, the mean μ and standard deviation σ. Strictly this distribution is a truncated Gaussian, however in practice we consider parameter ranges such that $p(m|\theta_{\rm F}) \ll 1$ at the boundaries. In contrast to the explicit differentiation between the true and observed SNR, Bayesian parameter estimation provides us with samples from the probability distribution of the true mass which are used directly as in Eq. (3.8), which eliminates the need to introduce a variable representing an observed mass.

In the background case, there are no selection effects, and we assume the noise characteristics are such that there is no correlation between the mass distribution and the SNR distribution. As a result, $p(\rho_i, \vec{m}_i|\theta_{\rm B}, \eta_i = {\rm B})$ decomposes as

$$p(\rho_i, \vec{m}_i|\theta_{\rm B}, \eta_i = {\rm B}) = p(\rho_i|\theta_{\rm B}, \eta_i = {\rm B}) p(\vec{m}_i|\theta_{\rm B}, \eta_i = {\rm B}). \tag{3.18}$$

Note that in realistic data the SNR distribution of background events may be strongly dependent on the mass (and other template parameters) [18] so this decomposition is not necessarily valid.

The expected rate and distribution of background events caused by instrumental noise can, in practice, be measured to high precision using techniques such as time-shifted analyses [93, 86] (see also [137]). For our artificial universe we have the freedom to choose the SNR and mass distributions, though this choice was informed by observed distributions in real data. We choose a power law with slope -12 in SNR; the mass posteriors are of constant width with their central values distributed uniformly between $m_{\rm min}$ and $m_{\rm max}$,

$$p(\rho|\eta = {\rm B}) \propto \rho^{-12}, \tag{3.19}$$

$$p(m|\eta = {\rm B}) \propto 1. \tag{3.20}$$

More realistic choices would include the effect of template bank density [138] and transient noise glitches [90, 92] on the distribution of noise triggers over

mass space. Note that our inference of the foreground mass distribution is expected to become more precise the more distinct the foreground and background are, especially in SNR. Here, both SNR distributions are falling power laws, however background drops off much more rapidly than foreground.

Finally we combine the mass distribution with Eqs. (3.14)-(3.15). Using (3.16) for the truncated power law we obtain:

$$\begin{aligned} &p(\rho, m|\alpha, m_{\rm low}, m_{\rm high}, \eta{=}\mathrm{F}) \\ &\quad\propto \int_{\rho_{\rm cutoff}(m)}^{\infty} \mathrm{d}\rho_{\rm true}\, \rho_{\rm true}^{-4}\, \chi_{\rm NC}(\rho; \lambda{=}\rho_{\rm true}, k{=}2) \\ &\qquad\times \begin{cases} m^{\alpha+3} & \text{if } m_{\rm low} < m < m_{\rm high} \\ 0 & \text{else} \end{cases} \end{aligned} \tag{3.21}$$

Using (3.17) for the Gaussian:

$$\begin{aligned} &p(\rho, m|\mu, \sigma, \eta{=}\mathrm{F}) \\ &\quad\propto \int_{\rho_{\rm cutoff}(m)}^{\infty} \mathrm{d}\rho_{\rm true}\, \rho_{\rm true}^{-4}\, \chi_{\rm NC}(\rho; \lambda{=}\rho_{\rm true}, k{=}2) \\ &\qquad\times \begin{cases} m^3 \mathcal{N}(m; \mu, \sigma) & \text{if } m_{\rm min} < m < m_{\rm max} \\ 0 & \text{else} \end{cases} \end{aligned} \tag{3.22}$$

The background model does not involve selection effects and yields

$$p(\rho, m|\eta{=}\mathrm{B}) \propto \rho^{-12}. \tag{3.23}$$

In general normalising these expressions requires an integral over ρ and m which can be difficult or computationally expensive. In our model this simplifies somewhat as the integrand $\rho_{\rm true}^{-4}\, \chi_{\rm NC}(\rho; \lambda = \rho_{\rm true}, k = 2)$ happens to assume values very close to zero for the $\rho_{\rm cutoff}$ values of $[0.25, 4]$ allowed by our prior mass range of $[5, 80]$, therefore we are able to approximate $\rho_{\rm cutoff} = 1$.

To generate the artificial datasets we draw a total number of foreground and background events from a Poisson distribution around the true values determined by the intrinsic rate. Each of those events corresponds necessarily to a local maximum of signal likelihood over time, mass, and, in general,

other parameters - we generally approximate this local maximum as a multivariate Gaussian distribution. For each foreground event we then draw the true mass and distance, from which we can uniquely determine the intrinsic SNR. We then simulate the impact of noise on the measurement of both SNR and mass by drawing a value from a non-central chi distribution around the true SNR to obtain the observed SNR, and drawing the maximum likelihood value from a normal distribution around the true mass with a width as determined by Eq. (3.13). We use a uniform in mass prior, therefore the posterior samples for the mass estimate are drawn from a Gaussian with the same width around the maximum likelihood value. For background events the observed SNR is drawn directly from a power-law as the background SNR distribution is determined empirically from the observed SNR values, while the posterior samples for the mass are drawn from a constant-width Gaussian around a central value drawn from a uniform distribution between $m_{\rm min}$ and $m_{\rm max}$ for each event.

Our method does not require us to make strong assumptions about the shape or width of the mass likelihood, however it is important to generate these artificial results carefully as negligence can have unexpected consequences. In practice, we have found the scaling and width of the posteriors to have little effect on our results when the posteriors are of smaller scale than the population.

3.4 Toy Model Results

We applied our method to a large number of realisations for each choice of foreground distribution, though the figures in the following section only show results for a single realisation. The results across realisations will be given in text only. The mass limits chosen for all toy model results[5] were $m_{\rm min} = 5$, $m_{\rm max} = 80$. The total expected number of events above an SNR of 8, the lowest threshold considered, is 1600, with 95% contamination due to background events. The chosen slope of -12 for the background SNR distribution is less steep than in typical LIGO-Virgo analyses for stellar-mass compact binary mergers; our choice exaggerates the transition region in which the chances of an event belonging to either foreground or background

[5]Since we do not claim a specific link to astrophysics in the toy model, the mass units are arbitrary.

are comparable.

To simulate the limitations due to the computational costs of the analysis we impose a SNR threshold on events, assessing its influence on our inferences by varying its value between 8 and 30. Most of these SNR values would typically be considered as sub-threshold, since an SNR of ≈ 13.7 is required for an event to have a $p_{\rm astro}$ value of 50%, and to reach $p_{\rm astro} = 99\%$ an SNR of ≈ 24 is needed. These numbers are meaningful only in relation to this simulation. The actual relationship between SNR and $p_{\rm astro}$ varies between detection pipelines as they typically use additional information in their detection statistics to reduce the significance of background events. The number of detectors used in the network, as well as their sensitivities and the actual characteristics of the foreground and background distributions will also have an effect. A more realistic application is given in section 3.5. Lastly, the free parameters in Eq. (3.12) and Eq. (3.13) are chosen such that an event of true mass $m = 30$ at a notional distance of 400 Mpc has a mass posterior with width $\sigma_{\rm PE} \approx 1$, and has an SNR of $\rho \approx 50$. The width of the mass posteriors of background events is set to the constant value of 3.2, typical of foreground events at the lowest SNR considered. In reality, the mass posterior distributions for background events are mass dependent and often irregular.

The priors chosen are flat in all hyper-parameters, with two exceptions: the width of Gaussian populations, where the prior was flat-in-log, and the expected number of astrophysical foreground events, where we used a Jeffreys prior:

$$\text{prior}(\sigma) \propto \frac{1}{\sigma}, \tag{3.24}$$

$$\text{prior}(N_{\rm F,exp}) \propto \frac{1}{\sqrt{N_{\rm F,exp}}}. \tag{3.25}$$

Parameter estimation was performed using the `emcee` [47] implementation of an Affine Invariant Markov chain Monte Carlo Ensemble sampler [46].

3.4.1 Power law distribution

The first population considered was the truncated power law, which was inspired by the idea that black hole masses may be distributed analogously to the initial mass function of their progenitor stars. We add parameters $m_{\rm min}$

and $m_{\max}$ to define the lower and upper limits of the power law distribution. This is motivated by the desire to determine whether there are gaps in the astrophysical black hole mass distribution: at the low end to compare with the apparent lower limit of black hole mass in X-ray binaries [139], and at the high end to determine the maximum mass above which a pair-instability supernova completely disrupts the star [140]. In our simulation we chose the power law slope to be -2.4, and the cut-off values to be 12 and 64.

Our primary results, the estimates of model parameters and their correlations, are shown in figure 3.1. These results use a SNR threshold 8, the lowest value for which we run our analysis, as we would expect this to yield the best possible parameter estimates. Notable features are the large spread of possible merger rate densities (abbreviated as "Rate") and their correlation with the lower mass cut-off. This is a consequence of the fact that the power law slope is effectively increased by 3 due to selection effects, thus detected events are described by a positive slope. The detection bias towards high mass means that fewer events are available to constrain the lower cut-off value, and low mass events which are observed tend to have lower SNR values. As the total rate is still dominated by low mass (and low-amplitude) events, the large uncertainty of the low mass cut-off yields a high uncertainty on the rate. The estimated fraction of foreground events in the sample of observed events couples linearly to the merger rate density, but is less significant than the lower mass cut-off.

To assess our method in the light of its main goal of avoiding bias while lowering the SNR threshold, a single analysis result is insufficient. Therefore, we analyse the same data with a range of different SNR thresholds to observe the change in the hyper-parameter estimates. Figure 3.2 shows the marginalised posteriors for the rate and the three population parameters as a function of SNR threshold. We can observe the posteriors growing wider as the SNR threshold is increased and information from fewer events is considered. The result from one single realisation is, however, not necessarily representative of the general behaviour. Combining the results from multiple realisation shows there is no noticeable bias regardless of the threshold chosen, and estimates improve as the threshold is lowered. Between SNR thresholds of 8 and 24, the width of the 90% credible intervals decreases on average by factors of 2.4 for the power-law slope, 1.3 and 1.6 for the lower and upper mass cut-offs respectively, and 1.7 for the log of the inferred

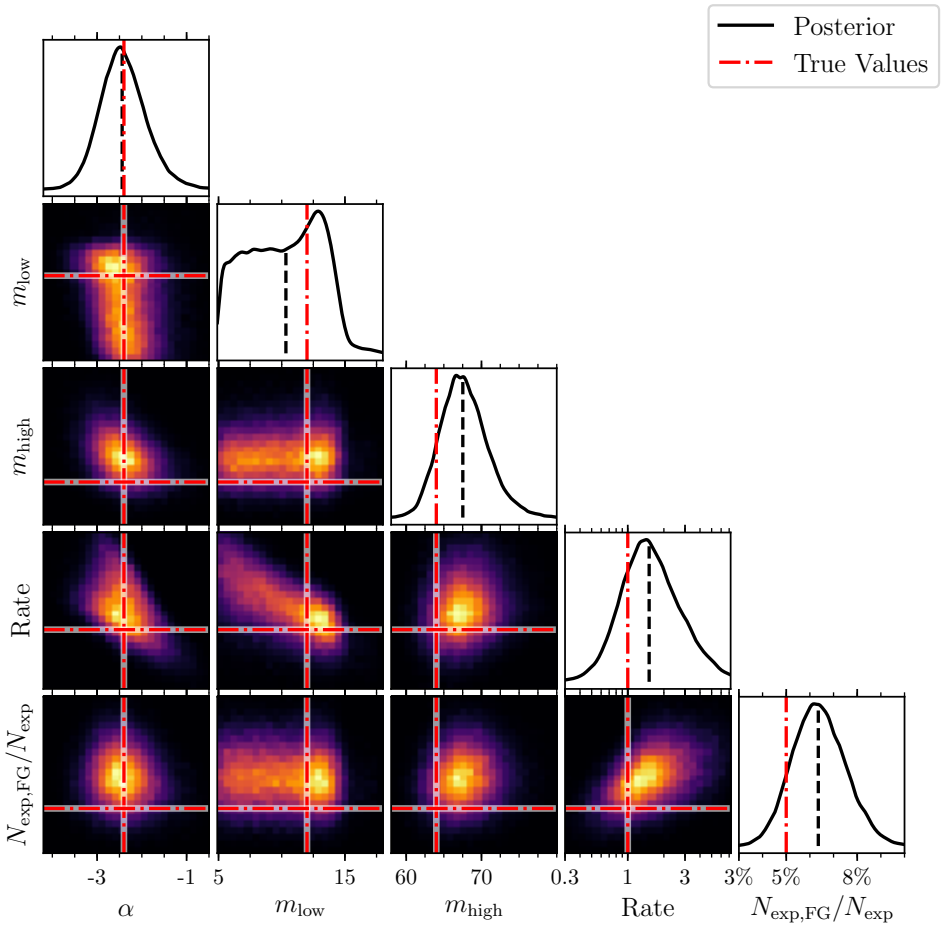

Figure 3.1: Parameter estimates for a single realisation of the toy model described in Sec 3.4.1. The foreground population model is a truncated power law with slope α, and cut-offs $m_{\rm min}$ and $m_{\rm max}$. The expected number of events above the SNR threshold of 8 are 1600, 5% of which are expected to be foreground events. The black lines show the kernel density estimate of the posterior (solid) and its median value (dashed). The red dash-dotted line indicates the true value for the underlying population.

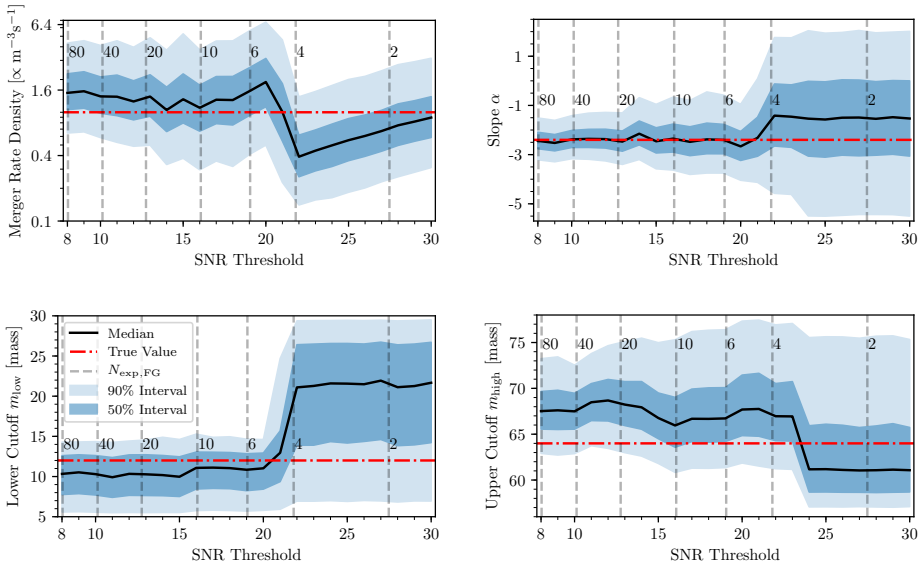

Figure 3.2: Confidence intervals for individual parameters of one realisation of the truncated power-law model (see Sec. 3.4.1), as a function of SNR threshold. The parameters shown are the inferred astrophysical merger rate (upper left) and power law slope (upper right), as well as the low (lower left) and high (lower right) mass cut-offs. The red dash-dotted line indicates the true value for the underlying population. Dashed grey lines indicate the expected number of foreground events at the given SNR threshold.

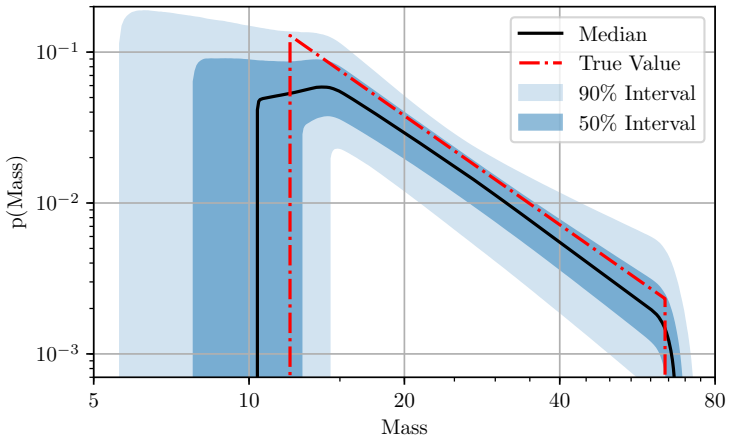

Figure 3.3: The inferred mass distribution of the foreground population using the truncated power law model and simulated data (see Sec. 3.4.1). The bands indicate the given percentiles in the probability density at any given mass across all posterior samples.

merger rate density.

Given the estimates of population parameters we can also compute an estimate of the underlying mass distribution, which we show in figure 3.3. Here we show the 50% and 90% confidence bands, as defined by computing the percentiles of $p(m|\theta)$ across all samples θ from the posterior for any given mass m. We observe that the true distribution is contained well within the credible interval and deviations are generally caused by an underestimated lower cut-off. In general there is a trade-off between expanding the bounds of the mass distribution to include additional events, and shrinking it to increase the PDF for highly significant events. The lower cut-off tends to have more freedom of movement as there are fewer high SNR events at low mass to constrain it.

As a final result for this population, figure 3.4 shows the estimated probability of any given event to have an astrophysical origin p_{astro}, and how it compares to a SNR-only estimate indicated by the black dash-dotted line. While this figure does not show quantitative results, we do observe that foreground events are largely located above the dash-dotted black line, indicating that our confidence in them being real has increased, while background events tend to be located below and are often on the $p_{\text{astro}} = 0$ line when their masses are outside the hard cut-offs of the truncated power law popula-

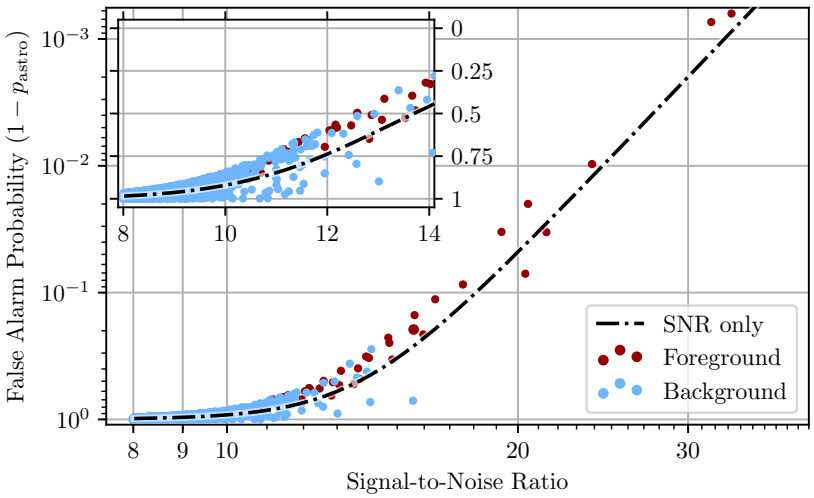

Figure 3.4: The probability that an event is caused by detector noise rather than being of astrophysical origin, $1 - p_{\mathrm{astro}}$, versus SNR. The foreground population model is a truncated power-law distribution. Blue and red dots represent foreground and background events respectively. The dash-dotted line shows the probability that would be inferred by a SNR based estimate, assuming the relative number of expected foreground and background events is known perfectly. The inset focuses on the region with background events and emphasises events which are unlikely to be astrophysical by using a linear scale.

tion. Thus we find, as expected, that the discrimination between signal and noise populations is improved with the incorporation of information about their mass distributions [138]. We determine that the percentage point difference between p_{astro} using our method and the SNR-only approach to be $\approx 2\%$, although this includes events with tiny absolute shifts due to them being very close to either 0% or 100% in the first place.

3.4.2 Gaussian distribution

The second core population considered is a simple Gaussian with a very small width. This population was chosen to test the inference on hard-to-infer parameters and to see the effect a very distinctive distribution has on the discriminating power of our method. We chose the width to be very narrow with a standard deviation of 1.6 around a mean mass of 27. The population is narrower than the individual posteriors, which have a typical width of $\approx 2-3$. Therefore we expect the population to require a relatively large number of events to resolve. How many events are needed to resolve these features generally depends on the population, in our case we find that $\approx 10-20$ events are needed to consistently constrain the population width to be smaller than individual posteriors. On the other hand, the discriminating power of using information from the mass estimates should be much greater than for a wide distribution like the truncated power law used in the previous section.

The parameter estimates for a single realisation are shown in figure 3.5. We observe that true width of the population σ is contained comfortably within the inferred posterior, though the uncertainty is rather large. It is generally overestimated slightly. Similarly the mean of the population is found well with an uncertainty comparable to the population width. The rate is constrained much better than in case of the truncated power law as this model lack the degeneracy between the rate and a poorly constrained population parameter. The lack of a strong correlation between a population parameter and the merger rate density also highlights its linear relation to the estimated number of foreground events contained within the sample.

When lowering the SNR threshold from 24 down to 8, the sizes of the 90% confidence intervals of the population parameters and merger rate density decrease by factors of 4.9, 2.0, 3.4 for the mean, the log of the width, and the log of the merger rate density, respectively. This is illustrated in figure 3.6 for

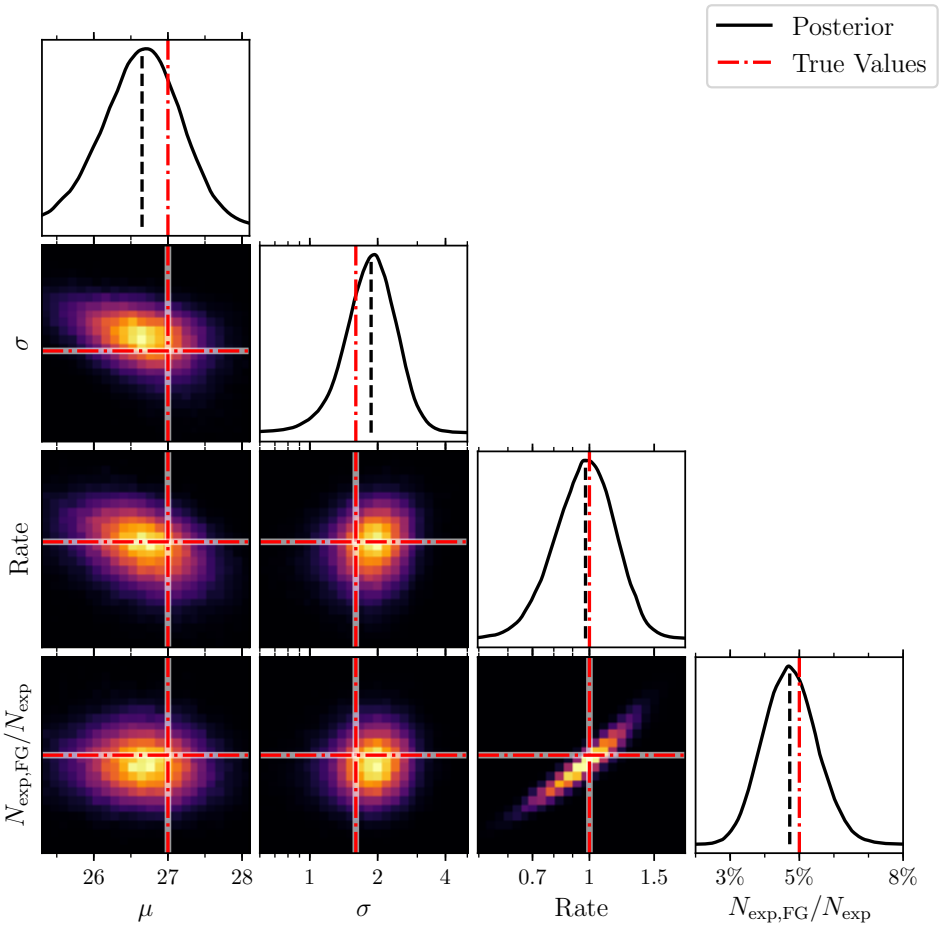

Figure 3.5: Parameter estimates for a single realisation of the Gaussian mass distribution described in Sec. 3.4.2. The foreground population model is a Gaussian with two free parameters, the mean μ and width σ. The true expected numbers of events above a SNR threshold 8 is 1600, 5% of which are expected to be foreground events. The black lines show the kernel density estimate of the posterior (solid) and its median value (dashed). The red dash-dotted lines indicates the true values for the underlying population.

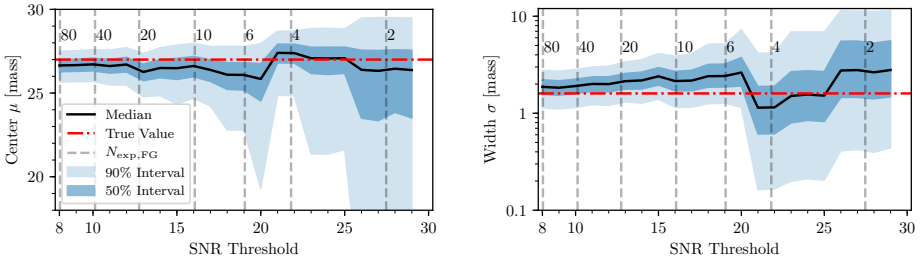

Figure 3.6: Confidence intervals for individual parameters of one realisation of the Gaussian population (see Sec. 3.4.2), as a function of SNR threshold. The parameters shown are the inferred mean (left) and the width of the distribution (right). The red dash-dotted line indicates the true value for the underlying population. Dashed grey lines indicate the expected number of foreground events at the given SNR threshold. The plot terminates at a SNR threshold of 29 as there are no events with SNR 30 or higher in this realisation.

one specific realisation. The uncertainty on inferred population parameters grows rapidly as the number of events decreases. The most noticable change is observed at SNR thresholds of 19, 20, and 21. This is caused by two events with SNRs between 19 and 20 whose removal decreases the number of events to 5, one of which is a low mass outlier. This outlier has an SNR slightly above 20, and is therefore removed when the SNR threshold is set to 21. The true mass distribution is well within the confidence interval shown in figure 3.7, though the true distribution is somewhat more narrow than inferred as seen previously in figure 3.5.

The comparison of the estimated p_{astro} as shown in figure 3.8 show how important the inclusion of masses is for this population. We can clearly identify the band of foreground events for which $1 - p_{\mathrm{astro}}$ is smaller by factors of a few up to 10 compared to the SNR-based estimate. In this specific realisation only six out of 76 foreground events lost any p_{astro}, across multiple simulations on average 97% of foreground events saw an increase in p_{astro}. In the case of background events $\approx 20\%$ saw an increase in their p_{astro} of up to 10%, though most are demoted and often down to effectively 0.

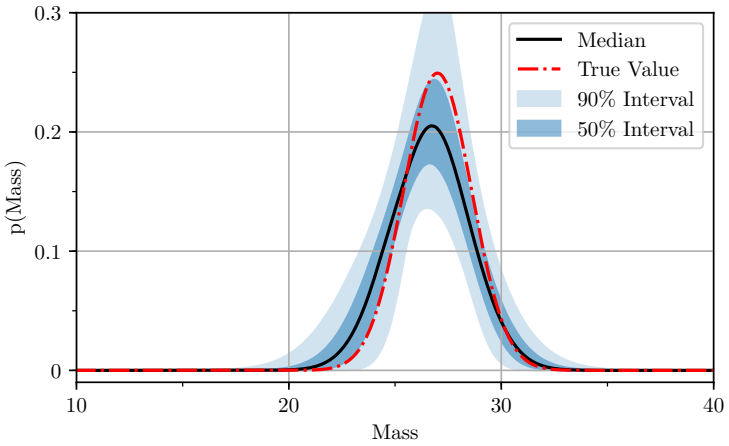

Figure 3.7: The inferred mass distribution of the foreground population using the Gaussian model and simulated dataset (see Sec. 3.4.2). The bands indicate the given percentiles in the probability density at any given mass across all posterior samples.

3.4.3 Incorrect models - No background component

Previous analyses of gravitational wave populations (such as the power law model used in [39]) use a high threshold to ensure a high probability that the events used are of astrophysical origin, in effect neglecting the possibility of background. Here we investigate the behaviour of our toy model with the background component disabled, corresponding to such a scenario. This shows the results one would obtain if simply fitting the foreground model to a contaminated dataset. The underlying population is a truncated power law identical to the one used in the first set of results presented in section 3.4.1.

The results are shown in Fig. 3.9, where we observe the inferred distribution to be very different from the true one when the lowest SNR threshold of 8 is used and the dataset is 95% polluted (left panel). The mass cut-offs are extended to the edges of the prior ranges to incorporate noise events at those values. The confidence interval includes the true value as long as the SNR threshold is sufficiently high since the number of background events is negligible, but trends towards -3 as the threshold is lowered. This is expected since the background dominates the low SNR region and has a power law slope of 0 in mass. This slope corresponds to an actual slope of -3 when

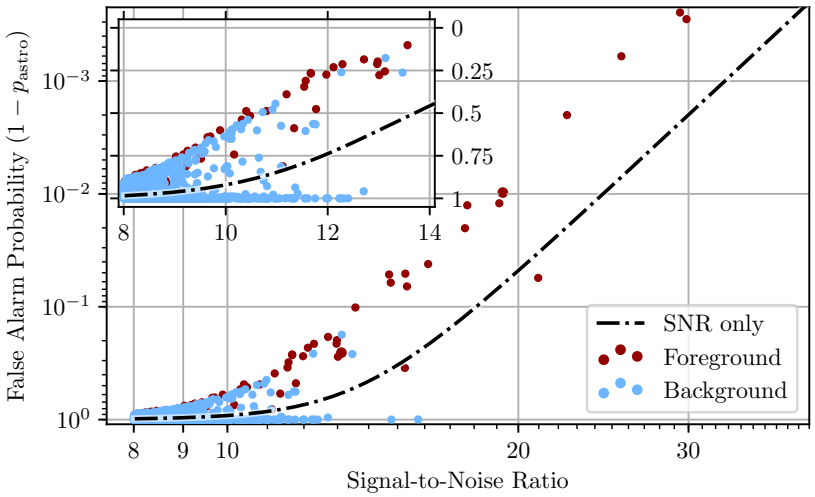

Figure 3.8: The probability of an event being caused by detector noise rather than being of astrophysical origin, $1 - p_{\mathrm{astro}}$, versus SNR. The foreground population model is a narrow Gaussian distribution (see Sec. 3.4.2). Blue and red dots represent foreground and background events respectively. The dash-dotted line shows the probability that would be inferred by a SNR based estimate, assuming the relative number of expected foreground and background events is known perfectly. The inset focuses on the region with background events and emphasises events which are unlikely to be astrophysical by using a linear scale.

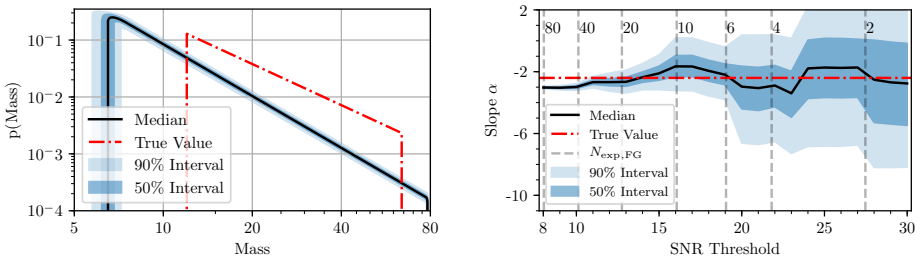

Figure 3.9: Results for the power law model when we neglect to account for contamination due to noise, as described in Sec. 3.4.3. Left: Inferred mass distribution at the lowest SNR threshold of 8, the PDF percentiles are calculated across the population posterior at any given mass. Right: Inferred power law slope as a function of SNR threshold. Vertical dashed lines indicate the expected number of true events above a particular SNR.

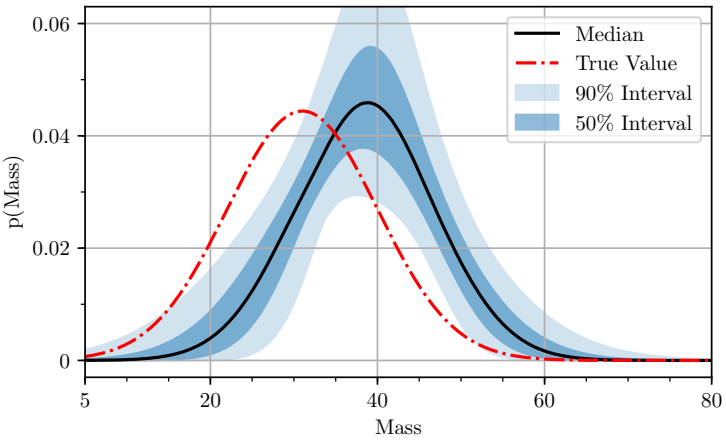

Figure 3.10: The inferred mass distribution for the Gaussian model without compensating for the mass dependence of selection effects (see Sec. 3.4.4), using the lowest SNR threshold of 8. The PDF percentiles are calculated across the population posterior at any given mass.

selection effects are considered. In the right panel we see the effect on the estimation of the power law slope as the threshold is varied. Once the SNR threshold reaches ≈ 12 the statistical uncertainty of the slope becomes large enough that the systematic bias is not noticeable.

3.4.4 Incorrect models - Neglected selection effects

The second kind of error we considered was to neglect to properly account for the mass dependence of selection effects. In case of a power law distribution this is trivial, as it simply adds +3 to the inferred value of the slope. Therefore we chose a Gaussian as the population, and we increased the width to 9 to highlight the impact of selection effects on the inferred population. Figure 3.10 shows that the selection effects effectively shift the distribution towards higher masses. This is a general feature as the m^3 term strongly favours high mass events in the observed set of events. Depending on the population this may also affect the width of the population, which happened to be a very minor effect in this case.

Together with the previous section 3.4.3 this illustrates that population inference can be made impossible even when the model matches the underlying distribution. Accounting for the presence of noise and selection effects

is essential for correct inference and to avoid bias when attempting to lower the SNR threshold.

3.5 Advanced LIGO engineering data simulation

After the successful tests using the simple toy model we also applied this method in a more realistic context, within an end-to-end analysis of simulated GW strain data where event candidates are identified by one of the compact binary detection pipelines currently used to search LIGO-Virgo data, `PyCBC` [141, 86, 142, 18]. We searched semi-realistic simulated data from the fourth advanced LIGO engineering run (ER4): LIGO Hanford Observatory (LHO) noise data were simulated as Gaussian noise using the LIGO design sensitivity noise spectrum with an angle-averaged range of $\sim 1600\,\mathrm{Mpc}$ for a $30 + 30\,\mathrm{M}_\odot$ black hole coalescence signal [29], while LIGO Livingston (LLO) data were derived from an instrumental channel monitoring the input laser power, recoloured to the same average target spectrum. The ER4 data contained a nontrivial population of high-amplitude noise transients, mostly arising from the LLO laser channel (LHO simulated data were also not entirely free of artefacts from data generation and transmission). However, these 'glitches' generally did not have similar morphology to binary merger signals, and did not give rise to a long-tailed background distribution, such as those occurring for some ranges of candidate parameters in recent Advanced detector data [94, 18, 27].

Simulated gravitational wave signals were added ("injected") to the noise data streams before they were stored and broadcast to the collaboration's computing grid. The injected population of binary black hole mergers was chosen to be uniform in both component masses between limits of 5 and 20 $\mathrm{M}_\odot$, and uniform in volume, with no cosmological (redshift) effects included. The `EOBNRv2HM` approximant tuned to numerical relativity [143] was used to simulate binary black hole mergers including non-dominant GW emission modes, for non-spinning binary components. The intended astrophysical rate corresponding to the injected merger signals was $5\,\mathrm{Gpc}^{-3}\mathrm{yr}^{-1}$.[6]

Our `PyCBC` search covered binary mergers of non-spinning components with masses between 3 and 50 $\mathrm{M}_\odot$; this range also defined the prior for

[6] Due to a software error the amplitude of injected signals was a factor 2 higher than intended, effectively simulating a true merger rate of $40\,\mathrm{Gpc}^{-3}\mathrm{yr}^{-1}$; however in the results presented here, we rescale our rate estimates to compensate for this error.

parameter estimation performed on each event using `LALinference` [41]. The search detection statistic for candidate events, ρ_c, is the quadrature sum of χ^2-reweighted SNRs $\hat{\rho}_{H,L}$ over single-detector events having consistent component masses and times of arrival between the two detectors [144, 107]. The number of events we chose to analyse is limited by computational cost; we impose a threshold $\rho_c > 8$ leaving us with 100 events in ≈ 37 days of LHO-LLO coincident observing time; 51 of these events correspond to known injected signals, with the remainder due to noise fluctuations.[7]

We first determine the rates of signal and noise events and the relative probabilities of signal vs. noise origin for each event [99, 85, 56], given only the ρ_c value of each event, models of the signal and noise event distributions over ρ_c, and an estimate of the total rate of noise events derived from time-shifted analyses [144, 142]. The result of this estimate is summarized in Fig. 3.11. We find 53 events with a signal probability p_{astro} above 50%, of which 47 have $p_{\mathrm{astro}} > 90\%$. This analysis is comparable to those used to estimate the rate and p_{astro} for binary black hole mergers in the first Advanced LIGO Observing period [85], and does not use information about the mass distributions of signal or noise events, besides the assumption that the signal population is contained within the analysis mass limits.

We now turn to our analysis, which estimates the rate and population model parameters simultaneously. The population model used here is a power law in each component mass, of the form

$$p(m_1, m_2|\theta_{\mathrm{F}}, \eta_i = \mathrm{F}) \propto \begin{cases} m_1^{\alpha} m_2^{\beta} & \text{if } m_{\mathrm{low}} < m_2 \leq m_1 < m_{\mathrm{high}} \\ 0 & \text{else} \end{cases} \tag{3.26}$$

where α and β are the two power law slopes, both with true values equal to 0. The mass cut-offs m_{low} and m_{high} are shared between both power laws, resulting in four free parameters in our population model. The selection effects were simulated numerically using the `LALsimulation` [31] implementation of the IMRPhenomPv2 waveform [21, 23] to implement the method described by Finn and Chernoff [32]. No particular cut-off was imposed on the asymmetric mass ratio q, though it is limited to $q > 0.25$ due to the range of allowed component masses.

As the background model, we used a power law fit to the distribution

[7] In reality we will not have access to an independent record listing all true signals!

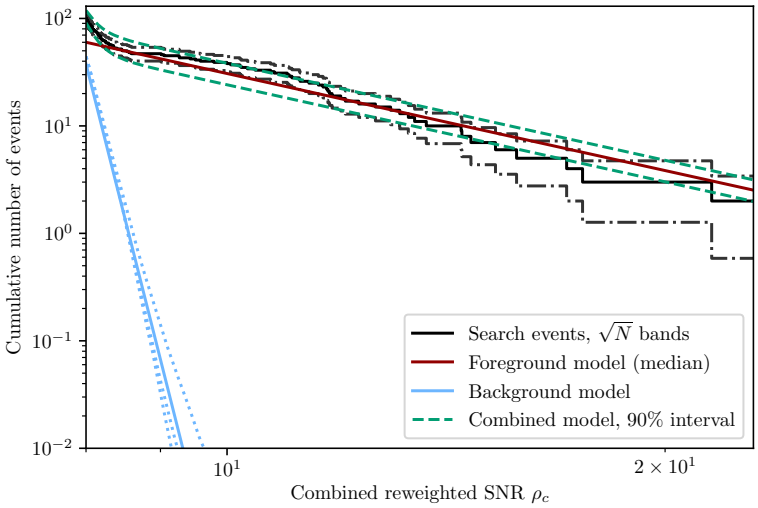

Figure 3.11: Cumulative number of (simulated) events detected in ≈ 37 days of LIGO ER4 engineering run data recolored to the 'ZDHP' design spectrum, vs. threshold search detection statistic ρ_c. Black steps indicate the search results, with $\pm\sqrt{N}$ bands indicating expected counting fluctuations. Dark red and light blue lines indicate power-law models of signal and noise distributions, respectively. Dotted light blue lines indicate empirical estimates of the noise distribution from each of 3 disjoint analysis periods, showing that the background model $p(\rho_c|\eta_i = \mathrm{B}) \propto \rho_c^{-54.8}$ is sufficiently accurate in the range of interest. Dark green dashed lines show the total expected number of events (signal+noise) as a 90% credible band.

of ρ_c values from the `PyCBC` time-shifted analysis, giving a slope of ≈ -54.8 for $\rho_c > 8$. In a separate step we constructed two dimensional fit to the distribution of component masses in the search results. We find empirically that the background distribution can be approximated as the product of a power-law in chirp mass and a exponential distribution in mass ratio. Using the masses as determined by the search is not strictly the correct approach, which would be to run the time-shifted data through the same LALInference analysis as used for the zero-lag events: this was computationally infeasible. Therefore the search results serve as a proxy for the optimal analysis.

We have found empirically that small changes to the background mass distribution do not have a strong effect on the result, which is expected given the dissimilarity of foreground and background distributions. (Note that real interferometer data containing chirp-like and blip-like features may be less forgiving in terms of the separability of foreground and background, as illustrated in Abbott et al. [18, 94].) We do not include any additional uncertainty on the background model rate or mass distribution.

The results of estimating the population parameters as displayed in figure 3.12 show that we successfully recover the true population parameters. The slopes are underestimated slightly which causes the inferred merger rate density to be elevated, although the true value is still encompassed. The mass cut-offs are found well with some tails to lower or higher masses for m_{low} and m_{high} respectively, since the uniform distribution of injections covers all regions of the mass range without major gaps and the cut-offs are shared between both component masses.

Since the slope of the background SNR distribution is much larger than the slope of the astrophysical foreground, the transition region where events may belong to either source category with comparable probabilities is quite small. This means there are few events which fall in between the region of certain background and certain foreground, limiting the gains that can be made by our method in this case. Nevertheless, the fact that the foreground mass distribution is quite distinct from the background causes a significant increase in p_{astro} relative to the ρ_c-based approach, as can be seen in figure 3.13.

As with the previous ranking statistic based analysis we can count the events found with a p_{astro} value above some given threshold. We find 56 events above a threshold of 50%, which is 3 events more than before. The

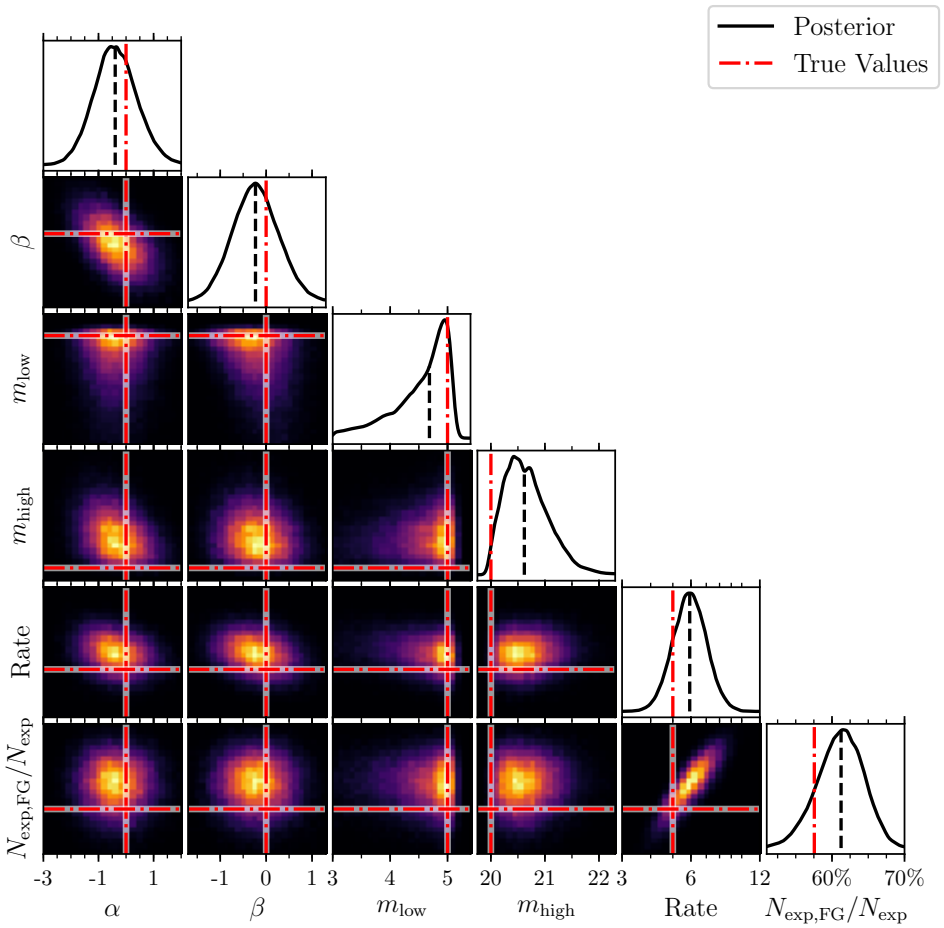

Figure 3.12: Parameter estimates for the ER4 dataset as summarized in Fig. 3.11 and Section 3.5. The foreground population model is power law in both component masses with separate slopes α and β and shared cut-offs $m_{\rm low}$ and $m_{\rm high}$. The cut-off masses and merger rate density are given in units of $M_\odot$ and $\mathrm{Gpc}^{-3}\mathrm{yr}^{-1}$ respectively. The black lines show the kernel density estimate of the posterior (solid) and its median value (dashed). The red dash-dotted line indicates the true value.

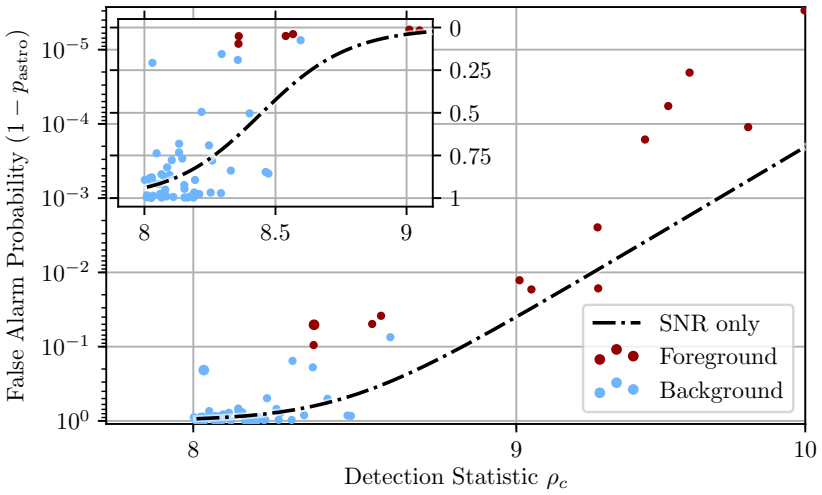

Figure 3.13: The probability of an event recovered from the ER4 data being caused by noise instead of corresponding to an astrophysical event, which is $1 - p_{\text{astro}}$. Blue and red dots represent foreground and background events respectively. The dash-dotted line shows the probability that would be inferred by a SNR based estimate, The inset focuses on the region with background events and emphasises events which are unlikely to be astrophysical by using a linear scale.

number of included events which we later identified as noise triggers rises from 4 to 5. When the threshold is set to $p_{\mathrm{astro}} > 90\%$, the number of events increases by 5 to 52, though it now includes one noise trigger. Thus, we find that applying our new method to realistic data not only reproduces the results obtained using established methods, but identifies additional foreground events. Since, in real Advanced detector data, the background distributions tend to be *less* steeply falling than those obtained here, we expect that the average (fractional) increase in number of signals found with high p_{astro} may be larger in the real-data case.

3.6 Outlook

In this work we have derived a new technique for simultaneous estimation of parameters defining the shape of two or more sub-populations and their expected contribution to the overall number of events. This technique allows us to extract information from formerly sub-threshold events without biasing the result due to uncertainties in classifying their origin. The method is agnostic to the specific choice of threshold, and lowering the threshold will improve the result by allowing information from events of an uncertain nature to be included.

Such improvements will eventually diminish in the regime where events have a high probability of background (noise) origin, though the point where further improvements become negligible depends on the specific characteristics of the two components. Figure 1 of Farr et al. [99] shows a simple example, using events for which only a single data value is measured, where improvements in foreground rate measurement become negligible in the background-dominated regime. In any case, digging deep into the background will not be detrimental; doing so can be especially worthwhile when the background model has some uncertainty, for instance following a specific functional form with partially unknown parameters. Additional events with low probability of astrophysical origin will reduce uncertainty on the background parameters, which in turn reduces the uncertainty on the foreground model.

The greatest gains over existing methods are found when there is a large number of events for which the source classification gives comparable probabilities for at least two categories, while the distributions in secondary

parameters are very distinct. This behaviour near the transition between populations is likely to be especially useful in the characterisation of weak event populations, such as unresolvable binary mergers at cosmological distances. This is of particular interest when determining whether the source population evolves with redshift. Conversely, gains are expected to be small when the primary source classification is very potent and population models are uncertain; in this case our method converges to that with a single population.

While such thresholded analyses that ignore the possible presence of background cannot be guaranteed free of systematic bias, the expected size of bias can be bounded by considering the rate of background events above threshold, as well as the degree of divergence between foreground and background distributions over the parameters of interest. Controlling the bias of a thresholded analysis thus still requires accurate background estimation. In particular, for the small number of high-significance events thus far detected by LIGO-Virgo, possible biases on population inference due to neglecting background contamination are expected to be well below statistical errors.

Furthermore, we find that selection effects must be included in the analysis to avoid systematic error of the population parameters. Our example study illustrates that this is particularly important for the mass distribution of binary black hole mergers.

We have successfully tested this new model on different binary merger mass distributions in an artificial universe, as well as to synthetic LIGO data from an engineering run. This demonstrates the feasibility of applying this method to existing and future LIGO-Virgo observing runs, which should allow a better joint determination of source event rates and distributions. Challenges in application to real data over a broad signal parameter space include adequately modelling the complex distribution of noise events over binary masses and spins [94, 18, 90]. The method itself is, however, not limited to the realm of gravitational wave astronomy, and can be useful whenever a set of data-points contains multiple populations.

Chapter 4

Conclusions

In this thesis we have considered multiple aspects of the gravitational wave astronomy of stellar-mass binary black-hole coalescences. We initially explored how well we can expect to be able to characterise GW150914-like gravitational wave signals using a number of different detector network configurations involving both existing detectors as well as those currently under construction. We found that the addition of a third detector is particularly important since it breaks the ring to which a GW signal is localised with two detectors, improving the localisation by factors larger than 20. This will be essential for attempts to identify the host galaxy of GW signals, be it to constrain potential associated EM emission, or to perform cosmological measurements [145]. For most other parameters the improvements were slightly below the naive estimate, which is mainly due to troublesome degeneracies.

Chapter 3 presents a new method for characterising the merger rate and the population of merging BBHs in the presence of noise and selection effects, although the method itself is general and not restricted to gravitational wave astronomy. After deriving the mathematical structure we created a number of test cases to confirm the constraints under which the model succeeds or fails, and if it fails, whether or not it does so as expected and when expected to. We found that the model produced correct results as long as it is close to the physical reality and improved continuously as the amount of data utilised was increased. It failed as expected if either the presence of noise or the existence of selection effects were not taken into account by our model.

The natural next step is to use the model developed in chapter 3 and

apply it to the real data taken by the LIGO-Virgo network during the first two observing periods (O1 & O2), or the currently ongoing third observing run. Doing this correctly involves a number of technical challenges due to the fact that we would like to include cosmological effects and use an injection campaign to assess the selection effects throughout the observing periods. The mixture model requires us to carefully normalise the different components of the likelihood, therefore any component of the likelihood which can not be integrated analytically poses a significant computational challenge.

Bibliography

[1] B. P. Abbott, R. Abbott, T. D. Abbott, et al. Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett.*, 116:061102, Feb 2016. doi: 10.1103/PhysRevLett.116.061102. URL `http://link.aps.org/doi/10.1103/PhysRevLett.116.061102`.

[2] J. M. Weisberg, J. H. Taylor, and L. A. Fowler. Gravitational waves from an orbiting pulsar. *Scientific American*, 245:74–82, October 1981. doi: 10.1038/scientificamerican1081-74.

[3] R. A. Hulse and J. H. Taylor. Discovery of a pulsar in a binary system. *Astrophys. J. Lett.*, 195:L51–L53, January 1975. doi: 10.1086/181708.

[4] S. M. Gaebel and J. Veitch. How would GW150914 look with future gravitational wave detector networks? *Classical and Quantum Gravity*, 34(17):174003, September 2017. doi: 10.1088/1361-6382/aa82d9.

[5] S. M. Gaebel, J. Veitch, T. Dent, and W. M. Farr. Digging the population of compact binary mergers out of the noise. *Mon. Notices Royal Astron. Soc.*, 484:4008–4023, April 2019. doi: 10.1093/mnras/stz225.

[6] A. Einstein. Die grundlage der allgemeinen relativitätstheorie. *Annalen der Physik*, 354(7):769–822, 1916. doi: 10.1002/andp.19163540702. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.19163540702`.

[7] Clifford M. Will. The confrontation between general relativity and experiment. *Living Reviews in Relativity*, 17(1):4, Jun 2014. ISSN 1433-8351. doi: 10.12942/lrr-2014-4. URL `https://doi.org/10.12942/lrr-2014-4`.

[8] The LIGO Scientific Collaboration, the Virgo Collaboration, B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, and et al. Tests of General Relativity with the Binary Black Hole Signals from the LIGO-Virgo Catalog GWTC-1. *arXiv e-prints*, art. arXiv:1903.04467, Mar 2019.

[9] Frank Watson Dyson, Arthur Stanley Eddington, and C. Davidson. Ix. a determination of the deflection of light by the sun's gravitational field, from observations made at the total eclipse of may 29, 1919. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 220(571-581):291–333, 1920. doi: 10.1098/rsta.1920.0009. URL `https://royalsocietypublishing.org/doi/abs/10.1098/rsta.1920.0009`.

[10] Event Horizon Telescope Collaboration, K. Akiyama, A. Alberdi, W. Alef, and et al. First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole. *Astrophys. J. Lett.*, 875: L1, April 2019. doi: 10.3847/2041-8213/ab0ec7.

[11] A. Einstein. Näherungsweise Integration der Feldgleichungen der Gravitation. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 688-696.*, 1916.

[12] C.W. Misner, K.S. Thorne, J.A. Wheeler, and D.I. Kaiser. *Gravitation*. Princeton University Press, 2017. ISBN 9781400889099. URL `https://books.google.co.uk/books?id=zAAuDwAAQBAJ`.

[13] S. W. Hawking. Black hole explosions? *Nature*, 248:30–31, March 1974. doi: 10.1038/248030a0.

[14] Frans Pretorius. Evolution of binary black-hole spacetimes. *Phys. Rev. Lett.*, 95:121101, Sep 2005. doi: 10.1103/PhysRevLett.95.121101. URL `https://link.aps.org/doi/10.1103/PhysRevLett.95.121101`.

[15] M. Campanelli, C. O. Lousto, P. Marronetti, and Y. Zlochower. Accurate evolutions of orbiting black-hole binaries without excision. *Phys. Rev. Lett.*, 96:111101, Mar 2006. doi: 10.1103/PhysRevLett.96.111101. URL `https://link.aps.org/doi/10.1103/PhysRevLett.96.111101`.

[16] John G. Baker, Joan Centrella, Dae-Il Choi, Michael Koppitz, and James van Meter. Gravitational-wave extraction from an inspiraling configuration of merging black holes. *Phys. Rev. Lett.*, 96:111102, Mar 2006. doi: 10.1103/PhysRevLett.96.111102. URL `https://link.aps.org/doi/10.1103/PhysRevLett.96.111102`.

[17] Yi Pan, Alessandra Buonanno, Andrea Taracchini, Lawrence E. Kidder, Abdul H. Mroué, Harald P. Pfeiffer, Mark A. Scheel, and Béla Szilágyi. Inspiral-merger-ringdown waveforms of spinning, precessing black-hole binaries in the effective-one-body formalism. *Phys. Rev. D*, 89(8):084006, 2014. doi: 10.1103/PhysRevD.89.084006.

[18] B. P. Abbott et al. GW150914: First results from the search for binary black hole coalescence with Advanced LIGO. *Phys. Rev.*, D93 (12):122003, 2016. doi: 10.1103/PhysRevD.93.122003.

[19] Andrea Taracchini, Alessandra Buonanno, Yi Pan, and et al. Effective-one-body model for black-hole binaries with generic mass ratios and spins. *Phys. Rev. D*, 89(6):061502, 2014. doi: 10.1103/PhysRevD.89.061502.

[20] Stanislav Babak, Andrea Taracchini, and Alessandra Buonanno. Validating the effective-one-body model of spinning, precessing binary black holes against numerical relativity. *Phys. Rev. D*, 95(2):024010, 2017. doi: 10.1103/PhysRevD.95.024010.

[21] Mark Hannam, Patricia Schmidt, Alejandro Bohé, Leïla Haegel, et al. Simple Model of Complete Precessing Black-Hole-Binary Gravitational Waveforms. *Phys. Rev. Lett.*, 113(15):151101, 2014. doi: 10.1103/PhysRevLett.113.151101.

[22] Sascha Husa, Sebastian Khan, Mark Hannam, Michael Pürrer, Frank Ohme, Xisco Jiménez Forteza, and Alejandro Bohé. Frequency-domain gravitational waves from nonprecessing black-hole binaries. I. New numerical waveforms and anatomy of the signal. *Phys. Rev. D*, 93(4): 044006, 2016. doi: 10.1103/PhysRevD.93.044006.

[23] Sebastian Khan, Sascha Husa, Mark Hannam, Frank Ohme, et al. Frequency-domain gravitational waves from non-precessing black-hole

binaries. II. A phenomenological model for the advanced detector era. *Phys. Rev. D*, 93:044007, 2016. doi: 10.1103/PhysRevD.93.044007.

[24] Stanislav Babak, Andrea Taracchini, and Alessandra Buonanno. Validating the effective-one-body model of spinning, precessing binary black holes against numerical relativity. *Phys. Rev. D*, 95(2):024010, Jan 2017. doi: 10.1103/PhysRevD.95.024010.

[25] B. P. Abbott, R. Abbott, T. D. Abbott, et al. Gw150914: The advanced ligo detectors in the era of first discoveries. *Phys. Rev. Lett.*, 116:131103, Mar 2016. doi: 10.1103/PhysRevLett.116.131103. URL `http://link.aps.org/doi/10.1103/PhysRevLett.116.131103`.

[26] Curt Cutler and Eanna E. Flanagan. Gravitational waves from merging compact binaries: How accurately can one extract the binary's parameters from the inspiral wave form? *Phys. Rev.*, D49:2658–2697, 1994. doi: 10.1103/PhysRevD.49.2658.

[27] L. K. Nuttall. Characterizing transient noise in the ligo detectors. *Phil. Trans. Roy. Soc. Lond.*, A376:20170286, 2018.

[28] B. P. Abbott, R. Abbott, T. D. Abbott, and et al. Gwtc-1: A gravitational-wave transient catalog of compact binary mergers observed by ligo and virgo during the first and second observing runs. *Phys. Rev. X*, 9:031040, Sep 2019. doi: 10.1103/PhysRevX.9.031040. URL `https://link.aps.org/doi/10.1103/PhysRevX.9.031040`.

[29] B. P. Abbott et al. Prospects for observing and localizing gravitational-wave transients with advanced ligo, advanced virgo and kagra. *Living Reviews in Relativity*, 21(1):3, Apr 2018. ISSN 1433-8351. doi: 10.1007/s41114-018-0012-9. URL `https://doi.org/10.1007/s41114-018-0012-9`.

[30] LIGO Scientific Collaboration. Gravitational wave interferometer noise calculator. `https://git.ligo.org/gwinc/pygwinc`, .

[31] LIGO Scientific Collaboration. Lsc algorithm library. `https://wiki.ligo.org/DASWG/LALSuite`, .

[32] Lee Samuel Finn and David F. Chernoff. Observing binary inspiral in gravitational radiation: One interferometer. *Phys. Rev. D*, 47:2198–2219, Mar 1993. doi: 10.1103/PhysRevD.47.2198. URL `https://link.aps.org/doi/10.1103/PhysRevD.47.2198`.

[33] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, and et al. Planck 2015 results. XIII. Cosmological parameters. *Astron. Astrophys.*, 594:A13, Sep 2016. doi: 10.1051/0004-6361/201525830.

[34] Astropy Collaboration, T. P. Robitaille, E. J. Tollerud, P. Greenfield, M. Droettboom, et al. Astropy: A community Python package for astronomy. *Astron. Astrophys.*, 558:A33, October 2013. doi: 10.1051/0004-6361/201322068.

[35] A. M. Price-Whelan, B. M. Sipőcz, H. M. Günther, (Astropy Contributors, et al. The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package. *Astron. J.*, 156:123, September 2018. doi: 10.3847/1538-3881/aabc4f.

[36] S. A. Usman, A. H. Nitz, I. W. Harry, et al. The PyCBC search for gravitational waves from compact binary coalescence. *ArXiv e-prints*, August 2015.

[37] Cody Messick, Kent Blackburn, Patrick Brady, et al. Analysis framework for the prompt discovery of compact binary mergers in gravitational-wave data. *Phys. Rev. D*, 95:042001, Feb 2017. doi: 10.1103/PhysRevD.95.042001. URL `https://link.aps.org/doi/10.1103/PhysRevD.95.042001`.

[38] S. Fairhurst. Source localization with an advanced gravitational wave detector network. *Classical and Quantum Gravity*, 28(10):105021, May 2011. doi: 10.1088/0264-9381/28/10/105021.

[39] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, and et al. Gw170814: A three-detector observation of gravitational waves from a binary black hole coalescence. *Phys. Rev. Lett.*, 119:141101, Oct 2017. doi: 10.1103/PhysRevLett.119.141101. URL `https://link.aps.org/doi/10.1103/PhysRevLett.119.141101`.

[40] Peter M. Lee. *Bayesian statistics : an introduction / Peter M. Lee.* 4th ed. edition, 2012.

[41] J. Veitch, V. Raymond, B. Farr, W. Farr, et al. Parameter estimation for compact binaries with ground-based gravitational-wave observations using the lalinference software library. *Phys. Rev. D*, 91: 042003, Feb 2015. doi: 10.1103/PhysRevD.91.042003. URL `https://link.aps.org/doi/10.1103/PhysRevD.91.042003`.

[42] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953. doi: 10.1063/1.1699114. URL `https://doi.org/10.1063/1.1699114`.

[43] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 04 1970. ISSN 0006-3444. doi: 10.1093/biomet/57.1.97. URL `https://dx.doi.org/10.1093/biomet/57.1.97`.

[44] John Skilling. Nested sampling. *AIP Conference Proceedings*, 735 (1):395–405, 2004. doi: 10.1063/1.1835238. URL `https://aip.scitation.org/doi/abs/10.1063/1.1835238`.

[45] John Skilling. Nested sampling for general bayesian computation. *Bayesian Anal.*, 1(4):833–859, 12 2006. doi: 10.1214/06-BA127. URL `https://doi.org/10.1214/06-BA127`.

[46] J. Goodman and J. Weare. Ensemble samplers with affine invariance. *Commun. Appl. Math. Comput. Sci.*, 65(5), 2010.

[47] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman. emcee: The mcmc hammer. *PASP*, 125:306–312, 2013. doi: 10.1086/670067.

[48] S. Stevenson, A. Vigna-Gómez, I. Mandel, J. W. Barrett, C. J. Neijssel, D. Perkins, and S. E. de Mink. Formation of the first three gravitational-wave observations through isolated binary evolution. *Nature Communications*, 8:14906, April 2017. doi: 10.1038/ncomms14906.

[49] J. W. Barrett, S. M. Gaebel, C. J. Neijssel, et al. Accuracy of inference on the physics of binary evolution from gravitational-wave observations. *Mon. Not. Roy. Astron. Soc.*, 477:4685–4695, July 2018. doi: 10.1093/mnras/sty908.

[50] J. Aasi et al. Advanced LIGO. *Class. Quant. Grav.*, 32:074001, 2015. doi: 10.1088/0264-9381/32/7/074001.

[51] B. P. Abbott et al. Properties of the Binary Black Hole Merger GW150914. *Phys. Rev. Lett.*, 116(24):241102, 2016. doi: 10.1103/PhysRevLett.116.241102.

[52] B. P. Abbott et al. Binary Black Hole Mergers in the first Advanced LIGO Observing Run. *Phys. Rev.*, X6(4):041015, 2016. doi: 10.1103/PhysRevX.6.041015.

[53] F. Acernese et al. Advanced Virgo: a second-generation interferometric gravitational wave detector. *Class. Quant. Grav.*, 32(2):024001, 2015. doi: 10.1088/0264-9381/32/2/024001.

[54] Yoichi Aso, Yuta Michimura, Kentaro Somiya, et al. Interferometer design of the kagra gravitational wave detector. *Phys. Rev. D*, 88: 043007, Aug 2013. doi: 10.1103/PhysRevD.88.043007. URL `http://link.aps.org/doi/10.1103/PhysRevD.88.043007`.

[55] Bala Iyer. "ligo-india, proposal of the consortium for indian initiative in gravitational-wave observations (indigo)". Technical Report LIGO-M1100296, LIGO Scientific Collaboration, 2011. URL `https://dcc.ligo.org/LIGO-M1100296/public`.

[56] B. P. Abbott et al. The Rate of Binary Black Hole Mergers Inferred from Advanced LIGO Observations Surrounding GW150914. *Astrophys. J.*, 833(1):L1, 2016. doi: 10.3847/2041-8205/833/1/L1.

[57] Christopher P. L. Berry et al. Parameter estimation for binary neutron-star coalescences with realistic noise during the Advanced LIGO era. *Astrophys. J.*, 804(2):114, 2015. doi: 10.1088/0004-637X/804/2/114.

[58] J. Veitch, I. Mandel, B. Aylott, et al. Estimating parameters of coalescing compact binaries with proposed advanced detector networks. *Phys. Rev. D*, 85:104045, May 2012. doi: 10.1103/PhysRevD.85.104045. URL `http://link.aps.org/doi/10.1103/PhysRevD.85.104045`.

[59] Samaya Nissanke, Jonathan Sievers, Neal Dalal, and Daniel Holz. Localizing compact binary inspirals on the sky using ground-based gravitational wave interferometers. *The Astrophysical Journal*, 739(2):99, 2011. URL http://stacks.iop.org/0004-637X/739/i=2/a=99.

[60] Leo P. Singer et al. The First Two Years of Electromagnetic Follow-Up with Advanced LIGO and Virgo. *Astrophys. J.*, 795(2):105, 2014. doi: 10.1088/0004-637X/795/2/105.

[61] C P L Berry, B Farr, W M Farr, et al. Early advanced ligo binary neutron-star sky localization and parameter estimation. *Journal of Physics: Conference Series*, 716(1):012031, 2016. URL http://stacks.iop.org/1742-6596/716/i=1/a=012031.

[62] Linqing Wen and Yanbei Chen. Geometrical expression for the angular resolution of a network of gravitational-wave detectors. *Phys. Rev. D*, 81:082001, Apr 2010. doi: 10.1103/PhysRevD.81.082001. URL http://link.aps.org/doi/10.1103/PhysRevD.81.082001.

[63] K. Grover, S. Fairhurst, B. F. Farr, et al. Comparison of gravitational wave detector network sky localization approximations. *Phys. Rev. D*, 89:042004, Feb 2014. doi: 10.1103/PhysRevD.89.042004. URL http://link.aps.org/doi/10.1103/PhysRevD.89.042004.

[64] Salvatore Vitale, Ryan Lynch, Vivien Raymond, Riccardo Sturani, John Veitch, and Philp Graff. Parameter estimation for heavy binary-black holes with networks of second-generation gravitational-wave detectors. 2016.

[65] Reed Essick, Salvatore Vitale, Erik Katsavounidis, Gabriele Vedovato, and Sergey Klimenko. Localization of short duration gravitational-wave transients with the early advanced ligo and virgo detectors. *The Astrophysical Journal*, 800(2):81, 2015. URL http://stacks.iop.org/0004-637X/800/i=2/a=81.

[66] Salvatore Vitale, Reed Essick, Erik Katsavounidis, Sergey Klimenko, and Gabriele Vedovato. On similarity of binary black hole gravitational-wave skymaps: to observe or to wait? 2016. doi: 10.1093/mnrasl/slw239.

[67] M. Pürrer. Frequency-domain reduced order models for gravitational waves from aligned-spin compact binaries. *Classical and Quantum Gravity*, 31(19):195010, October 2014. doi: 10.1088/0264-9381/31/19/195010.

[68] Andrea Taracchini, Alessandra Buonanno, Yi Pan, and et al. Effective-one-body model for black-hole binaries with generic mass ratios and spins. *Phys. Rev. D*, 89:061502, Mar 2014. doi: 10.1103/PhysRevD.89.061502. URL `http://link.aps.org/doi/10.1103/PhysRevD.89.061502`.

[69] B P Abbott, R Abbott, T D Abbott, et al. Characterization of transient noise in advanced ligo relevant to gravitational wave signal gw150914. *Classical and Quantum Gravity*, 33(13):134001, 2016. URL `http://stacks.iop.org/0264-9381/33/i=13/a=134001`.

[70] W. M. Farr. Computing credible areas and p-values for mcmc samples on the sky using a clustered-kernel-density estimate. `https://github.com/farr/skyarea`, 2013-2016.

[71] Hsin-Yu Chen, Reed Essick, Salvatore Vitale, Daniel E. Holz, and Erik Katsavounidis. Observational Selection Effects with Ground-based Gravitational Wave Detectors. *Astrophys. J.*, 835(1):31, 2017. doi: 10.3847/1538-4357/835/1/31.

[72] Warren G. Anderson, Patrick R. Brady, Jolien D. E. Creighton, and Eanna E. Flanagan. An Excess power statistic for detection of burst sources of gravitational radiation. *Phys. Rev.*, D63:042003, 2001. doi: 10.1103/PhysRevD.63.042003.

[73] Salvatore Vitale. Three observational differences for binary black holes detections with second and third generation gravitational-wave detectors. *Phys. Rev.*, D94(12):121501, 2016. doi: 10.1103/PhysRevD.94.121501.

[74] Salvatore Vitale and Matthew Evans. Parameter estimation for binary black holes with networks of third generation gravitational-wave detectors. 2016.

[75] Will M Farr, Ben Farr, and Tyson Littenberg. Modelling calibration errors in cbc waveforms. Technical Report LIGO-T1400682, LIGO Project, 2015. URL `https://dcc.ligo.org/P1500262/`.

[76] B. P. Abbott et al. Calibration of the Advanced LIGO detectors for the discovery of the binary black-hole merger GW150914. 2016.

[77] Walter Del Pozzo. Inference of the cosmological parameters from gravitational waves: application to second generation interferometers. *Phys. Rev.*, D86:043011, 2012. doi: 10.1103/PhysRevD.86.043011.

[78] Leo P. Singer et al. Going the Distance: Mapping Host Galaxies of LIGO and Virgo Sources in Three Dimensions Using Local Cosmography and Targeted Follow-up. *Astrophys. J.*, 829(1):L15, 2016. doi: 10.3847/2041-8205/829/1/L15.

[79] XiLong Fan, Christopher Messenger, and Ik Siong Heng. A bayesian approach to multi-messenger astronomy: Identification of gravitational-wave host galaxies. *The Astrophysical Journal*, 795(1): 43, November 2014.

[80] Bernard F. Schutz. Determining the Hubble Constant from Gravitational Wave Observations. *Nature*, 323:310–311, 1986. doi: 10.1038/323310a0.

[81] Stephen R. Taylor, Jonathan R. Gair, and Ilya Mandel. Hubble without the Hubble: Cosmology using advanced gravitational-wave detectors alone. *Phys. Rev.*, D85:023535, 2012. doi: 10.1103/PhysRevD.85.023535.

[82] Walter Del Pozzo, Tjonnie G. F. Li, and Chris Messenger. Cosmological inference using only gravitational wave observations of binary neutron stars. *Phys. Rev.*, D95(4):043502, 2017. doi: 10.1103/PhysRevD.95.043502.

[83] C. Messenger and J. Read. Measuring a cosmological distance-redshift relationship using only gravitational wave observations of binary neutron star coalescences. *Phys. Rev. Lett.*, 108:091101, 2012. doi: 10.1103/PhysRevLett.108.091101.

[84] Samaya Nissanke, Daniel E. Holz, Neal Dalal, Scott A. Hughes, Jonathan L. Sievers, and Christopher M. Hirata. Determining the Hubble constant from gravitational wave observations of merging compact binaries. 2013.

[85] B. P. Abbott et al. Binary Black Hole Mergers in the first Advanced LIGO Observing Run. *Phys. Rev. X*, 6(4):041015, 2016. doi: 10.1103/PhysRevX.6.041015.

[86] Samantha A. Usman, Alexander H. Nitz, Ian W. Harry, Christopher M. Biwer, et al. The PyCBC search for gravitational waves from compact binary coalescence. *Class. Quant. Grav.*, 33(21):215004, 2016. doi: 10.1088/0264-9381/33/21/215004.

[87] Kipp Cannon, Romain Cariou, Adrian Chapman, Mireia Crispin-Ortuzar, et al. Toward Early-Warning Detection of Gravitational Waves from Compact Binary Coalescence. *Astrophys. J.*, 748:136, 2012. doi: 10.1088/0004-637X/748/2/136.

[88] K. Cannon, C. Hanna, and J. Peoples. Likelihood-Ratio Ranking Statistic for Compact Binary Coalescence Candidates with Rate Estimation. *ArXiv e-prints*, April 2015.

[89] Cody Messick, Kent Blackburn, Patrick Brady, Patrick Brockill, et al. Analysis Framework for the Prompt Discovery of Compact Binary Mergers in Gravitational-wave Data. *Phys. Rev.*, D95(4):042001, 2017. doi: 10.1103/PhysRevD.95.042001.

[90] Alexander H. Nitz, Thomas Dent, Tito Dal Canton, Stephen Fairhurst, and Duncan A. Brown. Detecting binary compact-object mergers with gravitational waves: Understanding and Improving the sensitivity of the PyCBC search. *Astrophys. J.*, 849(2):118, 2017. doi: 10.3847/1538-4357/aa8f50.

[91] Bruce Allen. A chi**2 time-frequency discriminator for gravitational wave detection. *Phys. Rev. D*, 71:062001, 2005. doi: 10.1103/PhysRevD.71.062001.

[92] Alexander Harvey Nitz. Distinguishing short duration noise transients in LIGO data to improve the PyCBC search for gravitational waves

from high mass binary black hole mergers. *Class. Quant. Grav.*, 35 (3):035016, 2018. doi: 10.1088/1361-6382/aaa13d.

[93] C. Capano, T. Dent, C. Hanna, M. Hendry, et al. Systematic errors in estimation of gravitational-wave candidate significance. *Phys. Rev. D*, 96:082002, Oct 2017. doi: 10.1103/PhysRevD.96.082002. URL `https://link.aps.org/doi/10.1103/PhysRevD.96.082002`.

[94] B. P. Abbott et al. Characterization of transient noise in Advanced LIGO relevant to gravitational wave signal GW150914. *Class. Quant. Grav.*, 33(13):134001, 2016. doi: 10.1088/0264-9381/33/13/134001.

[95] Ely D. Kovetz, Ilias Cholis, Patrick C. Breysse, and Marc Kamionkowski. Black hole mass function from gravitational wave measurements. *Phys. Rev.*, D95(10):103010, 2017. doi: 10.1103/PhysRevD.95.103010.

[96] Stephen Fairhurst and Patrick Brady. Interpreting the results of searches for gravitational waves from coalescing binaries. *Class. Quant. Grav.*, 25:105002, 2008. doi: 10.1088/0264-9381/25/10/105002.

[97] Vaibhav Tiwari. Estimation of the Sensitive Volume for Gravitational-wave Source Populations Using Weighted Monte Carlo Integration. *ArXiv e-prints*, 2017.

[98] Ken K. Y. Ng, Salvatore Vitale, Aaron Zimmerman, Katerina Chatziioannou, Davide Gerosa, and Carl-Johan Haster. Gravitational-wave astrophysics with effective-spin measurements: asymmetries and selection biases. *ArXiv e-prints*, 2018.

[99] W. M. Farr, J. R. Gair, I. Mandel, and C. Cutler. Counting and confusion: Bayesian rate estimation with multiple populations. *Phys. Rev. D*, 91(2):023005, January 2015. doi: 10.1103/PhysRevD.91.023005.

[100] Rory Smith and Eric Thrane. Optimal Search for an Astrophysical Gravitational-Wave Background. *Phys. Rev.*, X8(2):021019, 2018. doi: 10.1103/PhysRevX.8.021019.

[101] T. G. F. Li, W. Del Pozzo, S. Vitale, et al. Towards a generic test of the strong field dynamics of general relativity using compact binary

coalescence. *Phys. Rev.*, D85:082003, 2012. doi: 10.1103/PhysRevD.85.082003.

[102] Walter Del Pozzo, Tjonnie G. F. Li, Michalis Agathos, et al. Demonstrating the feasibility of probing the neutron star equation of state with second-generation gravitational wave detectors. *Phys. Rev. Lett.*, 111(7):071101, 2013. doi: 10.1103/PhysRevLett.111.071101.

[103] C. Messenger and J. Veitch. Avoiding selection bias in gravitational wave astronomy. *New J. Phys.*, 15:053027, 2013. doi: 10.1088/1367-2630/15/5/053027.

[104] B. P. Abbott et al. Supplement: The Rate of Binary Black Hole Mergers Inferred from Advanced LIGO Observations Surrounding GW150914. *Astrophys. J. Suppl.*, 227(2):14, 2016. doi: 10.3847/0067-0049/227/2/14.

[105] Benjamin P. Abbott et al. Upper Limits on the Rates of Binary Neutron Star and Neutron Star–black Hole Mergers From Advanced Ligo's First Observing run. *Astrophys. J.*, 832(2):L21, 2016. doi: 10.3847/2041-8205/832/2/L21.

[106] Rahul Biswas, Patrick R. Brady, Jolien D. E. Creighton, and Stephen Fairhurst. The Loudest event statistic: General formulation, properties and applications. *Class. Quant. Grav.*, 26:175009, 2009. doi: 10.1088/0264-9381/26/17/175009,10.1088/0264-9381/30/7/079502. [Erratum: Class. Quant. Grav.30,079502(2013)].

[107] Abadie, J. et al. Search for Gravitational Waves from Low Mass Compact Binary Coalescence in LIGO's Sixth Science Run and Virgo's Science Runs 2 and 3. *Phys. Rev. D*, 85(8):082002, Apr 2012.

[108] R. O'Shaughnessy, B. Vaishnav, J. Healy, and D. Shoemaker. Intrinsic selection biases of ground-based gravitational wave searches for high-mass black hole-black hole mergers. *Phys. Rev. D*, 82:104006, Nov 2010. doi: 10.1103/PhysRevD.82.104006. URL `https://link.aps.org/doi/10.1103/PhysRevD.82.104006`.

[109] M. Dominik, E. Berti, R. O'Shaughnessy, I. Mandel, et al. Double Compact Objects III: Gravitational-wave Detection Rates. *Astrophys. J.*, 806:263, June 2015. doi: 10.1088/0004-637X/806/2/263.

[110] V. Kalogera, K. Belczynski, C. Kim, et al. Formation of double compact objects. *Physics Reports*, 442(1):75 – 108, 2007. ISSN 0370-1573. doi: https://doi.org/10.1016/j.physrep.2007.02.008. URL http://www.sciencedirect.com/science/article/pii/S0370157307000440. The Hans Bethe Centennial Volume 1906-2006.

[111] B. P. Abbott et al. Astrophysical implications of the binary black-hole merger gw150914. *The Astrophysical Journal Letters*, 818(2):L22, 2016. URL http://stacks.iop.org/2041-8205/818/i=2/a=L22.

[112] Ilya Mandel and Alison Farmer. Merging stellar-mass binary black holes. *ArXiv e-prints*, 2018.

[113] K. Belczynski et al. The Effect of Pair-Instability Mass Loss on Black Hole Mergers. *Astron. Astrophys.*, 594:A97, 2016. doi: 10.1051/0004-6361/201628980.

[114] K. Belczynski, J. Klencki, G. Meynet, C. L. Fryer, et al. The origin of low spin of black holes in LIGO/Virgo mergers. *ArXiv e-prints*, June 2017.

[115] K. Belczynski, D. E. Holz, T. Bulik, and R. O'Shaughnessy. The first gravitational-wave source from the isolated evolution of two stars in the 40-100 solar mass range. *Nature*, 534:512–515, June 2016. doi: 10.1038/nature18322.

[116] M. Spera, M. Mapelli, and A. Bressan. The mass spectrum of compact remnants from the PARSEC stellar evolution tracks. *Mon. Notices Royal Astron. Soc.*, 451:4086–4103, August 2015. doi: 10.1093/mnras/stv1161.

[117] Jim W. Barrett, Sebastian M. Gaebel, Coenraad J. Neijssel, Alejandro Vigna-Gómez, et al. Accuracy of inference on the physics of binary evolution from gravitational-wave observations. *Mon. Notices Royal Astron. Soc.*, 2017. doi: 10.1093/mnras/sty908.

[118] Michael Zevin, Chris Pankow, Carl L. Rodriguez, Laura Sampson, et al. Constraining Formation Models of Binary Black Holes with Gravitational-Wave Observations. *Astrophys. J.*, 846(1):82, 2017. doi: 10.3847/1538-4357/aa8408.

[119] Daniel Wysocki, Davide Gerosa, Richard O'Shaughnessy, Krzysztof Belczynski, et al. Explaining LIGO's observations via isolated binary evolution with natal kicks. *Phys. Rev.*, D97(4):043014, 2018. doi: 10.1103/PhysRevD.97.043014.

[120] Daniel Wysocki, Jacob Lange, and Richard O'Shaughnessy. Reconstructing phenomenological distributions of compact binaries via gravitational wave observations. *ArXiv e-prints*, 2018.

[121] C. Talbot and E. Thrane. Measuring the Binary Black Hole Mass Spectrum with an Astrophysically Motivated Parameterization. *Astrophys. J.*, 856:173, April 2018. doi: 10.3847/1538-4357/aab34c.

[122] S. R. Taylor and D. Gerosa. Mining Gravitational-wave Catalogs To Understand Binary Stellar Evolution: A New Hierarchical Bayesian Framework. *ArXiv e-prints*, June 2018.

[123] Ilya Mandel, Will M. Farr, Andrea Colonna, Simon Stevenson, et al. Model-independent inference on compact-binary observations. *Mon. Not. Roy. Astron. Soc.*, 465(3):3254–3260, 2017. doi: 10.1093/mnras/stw2883.

[124] R. O'Shaughnessy, D. Gerosa, and D. Wysocki. Inferences about Supernova Physics from Gravitational-Wave Measurements: GW151226 Spin Misalignment as an Indicator of Strong Black-Hole Natal Kicks. *Phys. Rev. Lett.*, 119(1):011101, July 2017. doi: 10.1103/PhysRevLett.119.011101.

[125] I. Mandel and R. O'Shaughnessy. Compact binary coalescences in the band of ground-based gravitational-wave detectors. *Classical and Quantum Gravity*, 27(11):114007, June 2010. doi: 10.1088/0264-9381/27/11/114007.

[126] D. Gerosa, M. Kesden, E. Berti, et al. Resonant-plane locking and spin alignment in stellar-mass black-hole binaries: A diagnostic of compact-binary formation. *Phys. Rev. D*, 87(10):104028, May 2013. doi: 10.1103/PhysRevD.87.104028.

[127] Simon Stevenson, Frank Ohme, and Stephen Fairhurst. Distinguishing compact binary population synthesis models using gravitational-wave

observations of coalescing binary black holes. *Astrophys. J.*, 810(1): 58, 2015. doi: 10.1088/0004-637X/810/1/58.

[128] Will M. Farr, Simon Stevenson, M. Coleman Miller, Ilya Mandel, et al. Distinguishing Spin-Aligned and Isotropic Black Hole Populations With Gravitational Waves. *Nature*, 548:426, 2017. doi: 10.1038/nature23453.

[129] Simon Stevenson, Christopher P. L. Berry, and Ilya Mandel. Hierarchical analysis of gravitational-wave measurements of binary black hole spin–orbit misalignments. *Mon. Not. Roy. Astron. Soc.*, 471(3): 2801–2811, 2017. doi: 10.1093/mnras/stx1764.

[130] Davide Gerosa and Emanuele Berti. Are merging black holes born from stellar collapse or previous mergers? *Phys. Rev.*, D95(12):124046, 2017. doi: 10.1103/PhysRevD.95.124046.

[131] Maya Fishbach, Daniel E. Holz, and Ben Farr. Are ligo's black holes made from smaller black holes? *Astrophys. J.*, 840(2):L24, 2017. doi: 10.3847/2041-8213/aa7045.

[132] Colm Talbot and Eric Thrane. Determining the population properties of spinning black holes. *Phys. Rev.*, D96(2):023012, 2017. doi: 10.1103/PhysRevD.96.023012.

[133] Salvatore Vitale, Ryan Lynch, Riccardo Sturani, and Philip Graff. Use of gravitational waves to probe the formation channels of compact binaries. *Classical and Quantum Gravity*, 34:03LT01, February 2017. doi: 10.1088/1361-6382/aa552e.

[134] Vaibhav Tiwari, Stephen Fairhurst, and Mark Hannam. Constraining black-hole spins with gravitational wave observations. *arXiv e-prints*, 2018.

[135] Manuela Campanelli, C. O. Lousto, and Y. Zlochower. Spinning-black-hole binaries: The orbital hang up. *Phys. Rev.*, D74:041501, 2006. doi: 10.1103/PhysRevD.74.041501.

[136] Ilya Mandel. Parameter estimation on gravitational waves from multiple coalescing binaries. *Phys. Rev.*, D81:084029, 2010. doi: 10.1103/PhysRevD.81.084029.

[137] Kipp Cannon, Chad Hanna, and Drew Keppel. Method to estimate the significance of coincident gravitational-wave observations from compact binary coalescence. *Phys. Rev. D*, 88(2):024025, 2013. doi: 10.1103/PhysRevD.88.024025.

[138] T. Dent and J. Veitch. Optimizing gravitational-wave searches for a population of coalescing binaries: Intrinsic parameters. *Phys. Rev.*, D89(6):062002, 2014. doi: 10.1103/PhysRevD.89.062002.

[139] Will M. Farr, Niharika Sravan, Andrew Cantrell, Laura Kreidberg, et al. The Mass Distribution of Stellar-Mass Black Holes. *Astrophys. J.*, 741:103, 2011. doi: 10.1088/0004-637X/741/2/103.

[140] Z. Barkat, G. Rakavy, and N. Sack. Dynamics of Supernova Explosion Resulting from Pair Formation. *Physical Review Letters*, 18:379–381, March 1967. doi: 10.1103/PhysRevLett.18.379.

[141] Tito Dal Canton et al. Implementing a search for aligned-spin neutron star-black hole systems with advanced ground based gravitational wave detectors. *Phys. Rev.*, D90(8):082004, 2014. doi: 10.1103/PhysRevD.90.082004.

[142] Alex Nitz, Ian Harry, Duncan Brown, Christopher M. Biwer, et al. gwastro/pycbc: Post-o2 release 12c, July 2018. URL `https://doi.org/10.5281/zenodo.1313589`.

[143] Yi Pan, Alessandra Buonanno, Michael Boyle, Luisa T. Buchman, et al. Inspiral-merger-ringdown multipolar waveforms of nonspinning black-hole binaries using the effective-one-body formalism. *Phys. Rev.*, D84:124052, 2011. doi: 10.1103/PhysRevD.84.124052.

[144] S. Babak, R. Biswas, P. R. Brady, D. A. Brown, et al. Searching for gravitational waves from binary coalescence. *Phys. Rev. D*, 87: 024033, Jan 2013. doi: 10.1103/PhysRevD.87.024033. URL `https://link.aps.org/doi/10.1103/PhysRevD.87.024033`.

[145] M. Soares-Santos, A. Palmese, W. Hartley, LIGO Scientific Collaboration, and Virgo Collaboration. First Measurement of the Hubble Constant from a Dark Standard Siren using the Dark Energy Survey Galaxies and the LIGO/Virgo Binary-Black-hole Merger GW170814.

Astrophys. J. Lett., 876(1):L7, May 2019. doi: 10.3847/2041-8213/ab14f1.

Appendix

The following paper is taken verbatim from *Mon. Notices Royal Astron. Soc.*. This paper was led by Jim W. Barrett [49], while I contributed the section describing selection effects, as well as the code to simulate selection effects in `COMPAS` [48] using `lalsimulation` [31].

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **477,** 4685–4695 (2018) doi:10.1093/mnras/sty908
Advance Access publication 2018 April 12

Accuracy of inference on the physics of binary evolution from gravitational-wave observations

Jim W. Barrett,[1]★ Sebastian M. Gaebel,[1] Coenraad J. Neijssel,[1] Alejandro Vigna-Gómez,[1] Simon Stevenson,[1,2] Christopher P. L. Berry,[1]★ Will M. Farr[1] and Ilya Mandel[1]

[1] *Institute of Gravitational-wave Astronomy and School of Physics and Astronomy, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK*
[2] *OzGrav, Swinburne University of Technology, Hawthorn, VIC 3122, Australia*

Accepted 2018 April 9. Received 2018 April 9; in original form 2017 November 17

ABSTRACT

The properties of the population of merging binary black holes encode some of the uncertain physics underlying the evolution of massive stars in binaries. The binary black hole merger rate and chirp-mass distribution are being measured by ground-based gravitational-wave detectors. We consider isolated binary evolution, and explore how accurately the physical model can be constrained with such observations by applying the Fisher information matrix to the merging black hole population simulated with the rapid binary-population synthesis code COMPAS. We investigate variations in four COMPAS parameters: common-envelope efficiency, kick-velocity dispersion and mass-loss rates during the luminous blue variable, and Wolf–Rayet stellar-evolutionary phases. We find that ∼1000 observations would constrain these model parameters to a fractional accuracy of a few per cent. Given the empirically determined binary black hole merger rate, we can expect gravitational-wave observations alone to place strong constraints on the physics of stellar and binary evolution within a few years. Our approach can be extended to use other observational data sets; combining observations at different evolutionary stages will lead to a better understanding of stellar and binary physics.

Key words: black hole physics – gravitational waves – stars: black holes – stars: evolution.

1 INTRODUCTION

Gravitational waves from binary black hole coalescences (Abbott et al. 2016a, 2017a,b,c) have recently been observed by the ground-based gravitational-wave detectors of the Advanced Laser Interferometer Gravitational-Wave Observatory (aLIGO; Aasi et al. 2015) and Advanced Virgo (Acernese et al. 2015). These observations provide a revolutionary insight into the properties of the population of binary black holes. The catalogue of detections will grow rapidly as the instruments continue to improve their sensitivity (Abbott et al. 2018). In this paper, we analyse how such a catalogue will make it possible to infer the physics of binary evolution by performing inference on parametrized population synthesis models.

A number of channels for the formation of binary black holes have been proposed (see e.g. Abbott et al. 2016c; Miller 2016; Mandel & Farmer 2017, for reviews). In this study, we assume that all merging binary black holes form through classical isolated binary evolution via a common-envelope phase (Postnov & Yungelson 2014; Belczynski et al. 2016). While all events observed to date are consistent with having formed through this channel (Eldridge et al. 2017; Stevenson et al. 2017a; Giacobbo, Mapelli & Spera 2018), a future analysis would need to hierarchically include the possibility of contributions from multiple channels (e.g. Stevenson, Berry & Mandel 2017b; Talbot & Thrane 2017; Zevin et al. 2017).

Previous efforts to explore how stellar and binary population synthesis models could be constrained with gravitational-wave observations (e.g. Bulik & Belczyński 2003; Bulik, Belczyński & Rudak 2004; Mandel & O'Shaughnessy 2010; Gerosa et al. 2014; Stevenson, Ohme & Fairhurst 2015) have typically focused on a discrete set of models, usually obtained by varying one evolutionary parameter at a time (e.g. Voss & Tauris 2003; Dominik et al. 2012; Mennekens & Vanbeveren 2014). In this paper, we consider the realistic scenario in which the astrophysical model is described by a multidimensional set of continuous parameters which may be strongly correlated. We ask how well we could constrain these parameters with a large observational data set.

The main tool we use to tackle this problem is the Fisher (information) matrix. Fundamentally, if we make an observation of a process, and we have a model for that process that depends on some parameters, then the Fisher matrix quantifies how much we can learn about

★ E-mail: compas@star.sr.bham.ac.uk (JWB); cplb@star.sr.bham.ac.uk (CPLB)

the parameters in our model from the observation we made. We derive an expression for the Fisher matrix for binary-population synthesis. We use this to quantify how much we can learn about the population parameters from observations of binary black holes using the current generation of ground-based gravitational-wave detectors. While we concentrate on gravitational-wave observations here, the method is applicable to other data sets, and the best constraints may come from combining multiple complementary observations.

We use Fisher matrices to demonstrate that it may be possible to precisely measure the population parameters in binary-population synthesis models with ∼1000 observations of binary black hole mergers. At the expected rate of gravitational-wave detections (Abbott et al. 2017a), this could be within a few years of the detectors reaching design sensitivity (∼2–3 yr at design sensitivity for our fiducial model); the observing schedule for gravitational-wave observatories is given in Abbott et al. (2018).

We first give an introduction to our binary population synthesis model in Section 2, together with a description of the model parameters we wish to infer using gravitational-wave observations. In Section 3, we demonstrate how we transform the raw outputs of our binary population synthesis model by considering observational selection effects and redshift- and metallicity-dependent star formation rates. In Section 4, we introduce the statistical tools used in this paper: (i) the likelihood function representing the probability of an observation given our model; (ii) a method for including measurement uncertainties in observations; and (iii) the Fisher matrix, which quantifies the sensitivity of our model to changes in its underlying parameters. The results of applying this methodology to binary population synthesis models are presented and discussed in Section 5, and we discuss our conclusions in Section 6.

2 POPULATION SYNTHESIS OF MASSIVE STELLAR BINARIES

Many of the details of binary evolution are currently uncertain (Postnov & Yungelson 2014; De Marco & Izzard 2017). Population synthesis models efficiently, albeit approximately, simulate the interactions of a large number of binaries in order to capture population wide behaviour and thoroughly explore the space of initial conditions. Uncertainties in the physics underlying isolated binary evolution are captured within population synthesis models through tunable parameters, which we call population parameters. In this paper we focus on four population parameters which have an impact on binary black hole formation. We use the rapid population synthesis code COMPAS.[1] This uses the stellar evolutionary models of Hurley, Pols & Tout (2000). Final black hole masses are calculated using the delayed model of Fryer et al. (2012). With the exception of the variations to the four population parameters we describe in Section 2.1, we employ the Stevenson et al. (2017a) fiducial model throughout this paper.

2.1 Population parameters

2.1.1 Supernova kick velocity

The asymmetric ejection of matter (Janka & Mueller 1994; Burrows & Hayes 1996; Janka 2013) or emission of neutrinos (Woosley 1987; Bisnovatyi-Kogan 1993; Socrates et al. 2005) during a supernova can provide a kick to the stellar remnant. This birth kick is on the order of hundreds of km s^{-1} for neutron stars (Hobbs et al. 2005a). The typical strength of supernova kicks imparted to black holes is not well constrained observationally (Wong et al. 2014; Mandel 2016; Repetto, Igoshev & Nelemans 2017), although they may be reduced relative to neutron star through the accretion of material falling back on to the stellar remnant (Fryer et al. 2012).

In COMPAS, the strength of supernova kicks is parametrized using the dispersion parameter for a three-dimensional Maxwell–Boltzmann distribution $\sigma_{\rm kick}$. A kick velocity $v_{\rm kick}$ is drawn from the distribution

$$P(v_{\rm kick}) = \sqrt{\frac{2}{\pi}}\,\frac{v_{\rm kick}^2}{\sigma_{\rm kick}^3}\exp\left(\frac{-v_{\rm kick}^2}{2\sigma_{\rm kick}^2}\right). \quad (1)$$

Alternative parametrizations for the supernova kick have been considered by Bray & Eldridge (2016), who did not find sufficient evidence to prefer them; here, we consider only continuous variations to model parameters, including the kick velocity in the Maxwell–Boltzmann distribution.

The kick is modified to account for mass fallback, so that the final kick imparted to the black hole is

$$v_{\rm kick}^{*} = (1 - f_{\rm fb})v_{\rm kick}, \quad (2)$$

where $f_{\rm fb}$ is the fraction of matter that falls back on to the black hole, calculated according to the delayed model of Fryer et al. (2012). For carbon–oxygen core masses greater than 11 M$_\odot$, $f_{\rm fb} = 1$ and so many heavy black holes receive no natal kick in this model (Belczynski et al. 2016; Stevenson et al. 2017a). Whilst observations of the proper motions of isolated Galactic pulsars (Hobbs et al. 2005b) suggest a value of $\sigma_{\rm kick} = 265$ km s^{-1}, we choose a fiducial $\sigma_{\rm kick} = 250$ km s^{-1} to match Stevenson et al. (2017a).

2.1.2 Common-envelope efficiency

When mass transfer is dynamically unstable and initially proceeds on the very short dynamical time-scale of the donor, a shared, non-corotating common envelope is formed around the donor core and the companion (Paczynski 1976). The details of the common-envelope phase are amongst the least well understood across all phases of isolated binary evolution (for a review, see Ivanova et al. 2013).

In COMPAS, the classical energy formalism (Webbink 1984) is employed to parametrize uncertainty in the physics of the common envelope. When a binary begins a common-envelope phase, each envelope is bound to its core, with a total binding energy approximated by

$$E_{\rm bind} = -G\left[\frac{M_1(M_1 - M_{\rm core,1})}{\lambda_{\rm CE,1}R_1} + \frac{M_2(M_2 - M_{\rm core,2})}{\lambda_{\rm CE,2}R_2}\right], \quad (3)$$

where G is Newton's constant, $M_{\rm core,\,(1,2)}$ are the core masses of the two stars, $M_{(1,2)}$ and $R_{(1,2)}$ are the stellar masses and radii, respectively, and $\lambda_{\rm CE\,(1,2)}$ are the corresponding stellar-structure parameters introduced by de Kool (1990) and are functions of star's evolutionary state (e.g. Dewi & Tauris 2000; Kruckow et al. 2016).

The loss of co-rotation between the orbit of the cores and the common envelope leads to energy dissipation which causes the cores to spiral in. Some of this lost orbital energy may be eventually used to eject the common envelope. The efficiency with which this transfer of energy occurs is uncertain, and is characterized by the free parameter $\alpha_{\rm CE}$. In order to determine the separation after the common-envelope phase, the classical energy formalism compares the binding energy of the envelope to the energy transferred from

[1]Further details and sample COMPAS simulations are available at www.sr.bham.ac.uk/compas/.

the orbit $\Delta E_{\rm orbit}$ so that

$$E_{\rm bind} = \alpha_{\rm CE} \Delta E_{\rm orbit} \, . \tag{4}$$

If the binary has sufficient orbital energy to completely expel the envelope, we consider this a successful common-envelope event. Unsuccessful ejections lead to a merger before a binary black hole system is formed. We follow Stevenson et al. (2017a) in assuming that common-envelope phases initiated by main sequence of Hertzsprung gap donors always lead to mergers (cf. the pessimistic model of Dominik et al. 2012).

The fiducial choices of the parameters in COMPAS are $\lambda_{\rm CE} = 0.1$ and $\alpha_{\rm CE} = 1.0$. We explicitly leave $\lambda_{\rm CE}$ fixed whilst making small perturbations to $\alpha_{\rm CE}$; however, this is an issue of labelling, since it is the product of these two free parameters which is ultimately of importance to the common-envelope physics (Dominik et al. 2012).

2.1.3 *Mass-loss multipliers*

Throughout their lives, stars lose mass through stellar winds. The wind mass-loss rate depends strongly on the star's luminosity and is generally highest for high-mass, high-metallicity stars. The dearth of observations of low-metallicity environments means that wind mass-loss rates are poorly constrained at low metallicities, and at high masses where stars are intrinsically rare. These are precisely the regimes where the progenitors of gravitational-wave sources are likely to form (Belczynski et al. 2016; Eldridge & Stanway 2016; Lamberts et al. 2016; Stevenson et al. 2017a; Giacobbo et al. 2018).

COMPAS employs the approximate wind mass-loss prescriptions detailed in Belczynski et al. (2010). For hot O/B-stars, we employ the wind mass-loss prescription of Vink, de Koter & Lamers (2001). Our Wolf–Rayet wind mass-loss rates come from Hamann & Koesterke (1998). For other phases the mass-loss prescriptions from Hurley et al. (2000) are used. Uncertainty in mass-loss rates can have a significant impact on stellar evolution; for example, Renzo et al. (2017) find that there is an ∼50 per cent uncertainty in the mapping between initial and final masses when considering different mass-loss prescriptions when modelling solar-metallicity, non-rotating, single stars, with initial masses between 15 and 35 $\rm M_\odot$. There are particular phases of stellar evolution where the mass-loss rates lack strong constraints by observations. We parametrize the mass-loss rates in two of these phases with tunable population parameters.

During the luminous blue variable (LBV) phase (Humphreys & Davidson 1994), extremely massive stars undergo a relatively short episode of rapid mass-loss which strongly impacts the binary's future evolutionary trajectory (e.g. Mennekens & Vanbeveren 2014); observational constraints on the physics of LBV stars are currently uncertain (Smith 2017).[2] Following Belczynski et al. (2010), we parametrize this rate in terms of a multiplicative factor $f_{\rm LBV}$ used to modify the basic prescription, so that the rate becomes

$$\dot{M}_{\rm LBV} = f_{\rm LBV} \times 10^{-4} \ {\rm M_\odot \, yr^{-1}}; \tag{5}$$

our fiducial value for this factor is $f_{\rm LBV} = 1.5$ (Belczynski et al. 2010).

During the Wolf–Rayet phase, stars have lost their hydrogen envelopes and have high but relatively poorly constrained mass-loss rates (Crowther 2007). We use a multiplicative constant $f_{\rm WR}$ to modify the base rate:

$$\dot{M}_{\rm WR} = f_{\rm WR} \left(\frac{L}{\rm L_\odot} \right)^{1.5} \left(\frac{Z}{\rm Z_\odot} \right)^{m} \times 10^{-13} \ {\rm M_\odot \, yr^{-1}}, \tag{6}$$

where L is the stellar luminosity, Z is the metallicity, $Z_\odot = 0.02$ is approximately the bulk metallicity of our Sun, and $m = 0.86$ is an empirically determined scaling factor (Vink & de Koter 2005; Belczynski et al. 2010). The fiducial choice for this population parameter is $f_{\rm WR} = 1.0$. We use the same mass-loss prescription for all Wolf–Rayet subtypes (Belczynski et al. 2010), as the Hurley et al. (2000) evolutionary tracks do not distinguish between them. Recent investigations of mass-loss for Wolf–Rayet stars of varying composition include McClelland & Eldridge (2016), Tramper, Sana & de Koter (2016), and Yoon (2017).

3 MODEL PREDICTIONS

In this paper we evaluate the impact of the tunable parameters described above on the rate of detections and the measured chirp-mass distribution of binary black holes. The chirp mass $\mathcal{M}$ is a particular combination of the component masses M_1, M_2 which is measured well from the gravitational-wave frequency evolution during the binary inspiral (Cutler & Flanagan 1994; Abbott et al. 2016b),

$$\mathcal{M} = \frac{(M_1 M_2)^{3/5}}{(M_1 + M_2)^{1/5}}. \tag{7}$$

The chirp mass is just one of the parameters measurable through gravitational waves, other observables such as component masses, spins, and the distance to the source can also be inferred (Abbott et al. 2016b). For simplicity, we have chosen to focus on chirp mass since it is the best measured. This is a conservative approach, as we have neglected information about other parameters; however, the methods presented here are easily extendible to include other observables.

In order to represent the distribution of chirp masses produced by the population synthesis model, we chose to bin our systems by chirp mass. Throughout this paper, we use 30 bins of equal width, ranging from the lowest to the highest chirp masses present in our data set. The number of bins is determined by the scale length of variability in the chirp-mass distribution and the chirp-mass measurement uncertainty discussed below; the results are insensitive to halving the number of bins.

The raw output of a population synthesis model is a list of the initial conditions and final outcomes of all the binaries simulated. In order to compare this output to astronomical observations, it is necessary to process the data further, in order to account for the history of star formation in the Universe and the observational selection effects. We describe this processing below.

3.1 Cosmic history

In order to focus our computation on black hole progenitors, we only simulate systems with primary masses between $7 \, {\rm M_\odot} < M_1 < 100 \, {\rm M_\odot}$. We assume that all stars are in binaries with primary masses ranging between 0.01 and 150 $\rm M_\odot$ following the initial mass function of Kroupa (2001) with a flat mass-ratio distribution (Sana et al. 2012). At formation, binaries are assumed to have a uniform-in-the-logarithm distribution of orbital separations (Öpik 1924; Abt 1983) and zero orbital eccentricity; for more detailed studies of mass-ratio and orbital distributions, see Duchêne

[2] As in Hurley et al. (2000), we assume stars are in an LBV-like phase if their luminosity and radius satisfy $L > 6 \times 10^5 \, {\rm L_\odot}$ and $(R/{\rm R_\odot})(L/{\rm L_\odot})^{1/2} > 10^5$.

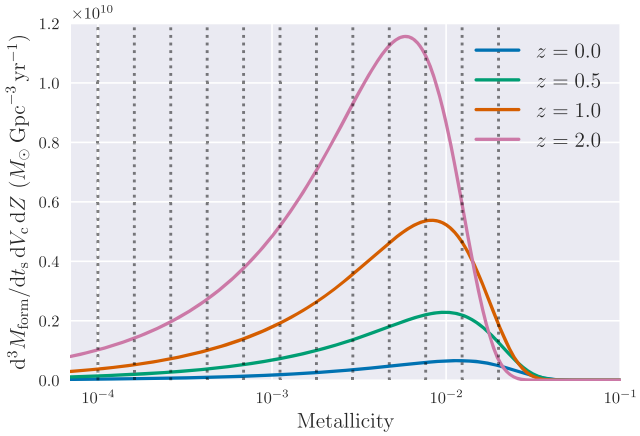

Figure 1. The metallicity-specific star formation rate as a function of metallicity at three different redshifts, using the star-formation-rate distribution of Madau & Dickinson (2014) and the metallicity distribution of Langer & Norman (2006). The vertical dashed lines indicate the metallicities at which we undertook simulations for this study. Metallicities above $Z_\odot = 0.02$ contribute negligibly to the binary black hole merger rate.

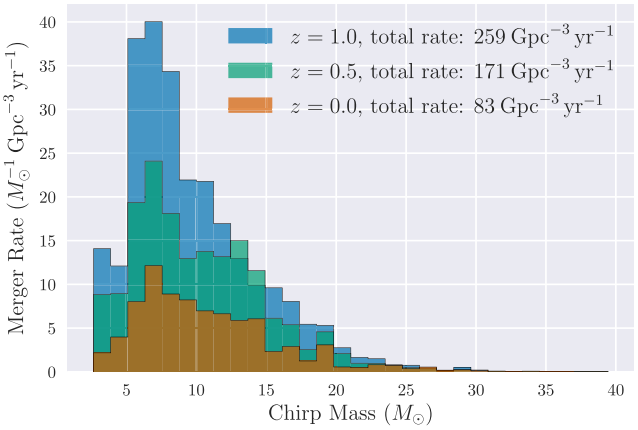

Figure 2. The binary black hole merger rate predicted by the COMPAS fiducial model at three different redshifts, taking into account the cosmic evolution of the metallicity-specific star formation rate. For comparison, the total inferred merger rate density from gravitational-wave observations is 12–213 ${\rm Gpc}^{-3}\,{\rm yr}^{-1}$ (Abbott et al. 2017a).

& Kraus (2013) and Moe & Di Stefano (2017). COMPAS simulations produce a rate of binary black hole formation per unit star formation mass $M_{\rm form}$,

$$\mathcal{R}_{\rm COMPAS} = \frac{{\rm d}^3 N_{\rm form}}{{\rm d}M_{\rm form}\,{\rm d}\tau_{\rm delay}\,{\rm d}\mathcal{M}}, \tag{8}$$

where $\tau_{\rm delay}$ is the delay time, defined as the time from the birth of a binary to its coalescence (Peters 1964). To compute the total rate of binary black hole mergers per unit comoving volume per unit time, we need to convolve the COMPAS formation rate with the amount of metallicity-specific star formation per unit volume per unit time at the birth of the binaries. Delay times can range from a few Myr to Gyr, and observations show that both the metallicity and star formation rates in galaxies evolve significantly over these time-scales (Madau & Dickinson 2014). We use the star formation rate distribution of Madau & Dickinson (2014) and the metallicity distribution of Langer & Norman (2006). Other distributions have been suggested (e.g. Savaglio et al. 2005; Vangioni et al. 2015; Ma et al. 2016), and the cosmic history of metallicity evolution adds an additional source of uncertainty to our model predictions. Future studies could consider how metallicity evolution could be included with the other model parameters and inferred from binary observations. In Fig. 1 we provide an illustration of the metallicity-specific star formation rate at redshifts $z = 0.5$, 1, and 1.5, and also indicate metallicities at which we performed simulations for this study. We use these to translate the star formation rate into the merger rate at redshift z

$$\frac{{\rm d}^3 N_{\rm merge}}{{\rm d}t_s {\rm d}V_c {\rm d}\mathcal{M}}(z) = \int {\rm d}Z \int {\rm d}\tau_{\rm delay} \left[\frac{{\rm d}^3 N_{\rm form}}{{\rm d}M_{\rm form}\,{\rm d}\tau_{\rm delay}\,{\rm d}\mathcal{M}}(Z) \times \frac{{\rm d}^3 M_{\rm form}}{{\rm d}t_s\,{\rm d}V_c\,{\rm d}Z}(Z, t_{\rm form} = t_{\rm merge}(z) - \tau_{\rm delay}) \right], \tag{9}$$

where t_s is the time measured in the frame of reference of the merger, V_c is the comoving volume and we use cosmological parameters from Planck Collaboration XIII (2016). Fig. 2 shows the local merger rate at three different redshifts after accounting for changes in star formation rate and cosmology.

3.2 Selection effects

Gravitational-wave detectors are not equally sensitive to every source. The distance to the source, its orientation and position relative to the detectors, as well as the physical characteristics of the source all affect how likely it is that the system would be detectable. The detectability of a signal depends upon its signal-to-noise ratio (SNR). The SNR in a single detector is defined as (Finn 1992)

$${\rm SNR}^2 = \langle h|h\rangle = 4\Re \int_{f_{\rm min}}^{f_{\rm max}} {\rm d}f\, \frac{h^*(f)h(f)}{S(f)}, \tag{10}$$

where $h(f)$ is the waveform measured by the detector, $S(f)$ is the one-sided noise power spectral density, and $f_{\rm min}$ and $f_{\rm max}$ are the limits of the frequency range considered.

For simplicity, we assume that signals are detected if their single-detector SNR exceeds a threshold value of 8 (Abbott et al. 2018). To model the waveforms, we use the IMRPhenomPv2 (Hannam et al. 2014; Husa et al. 2016; Khan et al. 2016) and SEOBNRv3 (Pan et al. 2014; Babak, Taracchini & Buonanno 2017) approximants;[3] these include the inspiral, merger, and ringdown phases of a binary black hole coalescence, and allow for precession of the black hole spins. We incorporate the effects of cosmological redshift, which manifest as an apparent increase in the system masses, $M_{\rm obs} = (1+z)M_s$ (Krolak & Schutz 1987; Holz & Hughes 2005). We assume a detector sensitivity equal to aLIGO in its design configuration (Aasi et al. 2015; Abbott et al. 2018).

We optimize our computations, reducing the number of waveform calculations required, by exploiting the fact that the parameters extrinsic to the gravitational-wave source, such as its position and orientation, only couple into the overall amplitude of the wave via

$$\mathcal{A} \propto \frac{1}{D_L}\sqrt{F_+^2(1+\cos^2 i)^2 + 4F_\times^2 \cos^2 i}, \tag{11}$$

$$F_+ \equiv \frac{1}{2}\cos(2\psi)[1+\cos^2(\theta)]\cos(2\phi) - \sin(2\psi)\cos(\theta)\sin(2\phi), \tag{12}$$

[3]We use the implementations publicly available in the LAL suite software package wiki.ligo.org/DASWG/LALSuite.

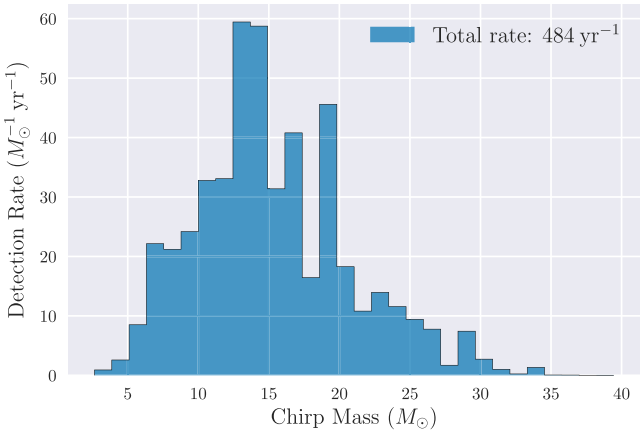

Figure 3. The rate and chirp-mass distribution of the binary black hole coalescences we expect aLIGO to observe at design sensitivity, taking into account cosmic history and selection effects, for the COMPAS fiducial model as described in Stevenson et al. (2017a). The detection rate is per unit observing time.

$$F_\times \equiv \frac{1}{2}\sin(2\psi)[1+\cos^2(\theta)]\cos(2\phi) + 2\cos(2\psi)\cos(\theta)\sin(2\phi), \quad (13)$$

where $\mathcal{A}$, D_L, i, ψ, θ, and φ are the gravitational-wave amplitude, luminosity distance, inclination, polarization, and polar and azimuthal angles of the source location in the detector frame, respectively (Krolak & Schutz 1987; Cutler & Flanagan 1994). Therefore, we need only compute the phase evolution for a given combination of intrinsic binary parameters, such as masses, once, and then marginalize over the extrinsic parameters (with the exception of D_L) as described in Finn & Chernoff (1993).

For a system with a given (M_1, M_2, D_L), we determine the fraction of extrinsic parameter realizations for which the observed SNR passes our threshold, and label this as our detection probability P_det.

We can use this detection probability to transform the merge rate given in equation (9) into a rate of detections. Integrating over the merger redshift gives the total detection rate

$$\frac{\mathrm{d}N_\mathrm{obs}}{\mathrm{d}t_\mathrm{obs}\,\mathrm{d}\mathcal{M}} = \int \mathrm{d}z \left[\frac{\mathrm{d}^3 N_\mathrm{merge}}{\mathrm{d}t_\mathrm{s}\,\mathrm{d}V_\mathrm{c}\,\mathrm{d}\mathcal{M}}\frac{\mathrm{d}V_\mathrm{c}}{\mathrm{d}z}\frac{\mathrm{d}t_\mathrm{s}}{\mathrm{d}t_\mathrm{obs}}P_\mathrm{det}\right], \quad (14)$$

where t_s is time in the source frame and $t_\mathrm{obs} = (1+z)t_\mathrm{s}$ is time in the observer's frame.

Fig. 3 shows the rate and chirp-mass distribution of binary black hole mergers detected at aLIGO design sensitivity. The mass distribution is shifted to higher masses relative to the intrinsic merger rate plotted in Fig. 2 because selection effects favour heavier systems which emit louder gravitational-wave signals. Some of the sharp features in this plot are the consequence of simulating systems on a discrete grid of metallicities (cf. Dominik et al. 2013). LBV winds tend to reduce high-mass stars to a narrow, metallicity-dependent range of black hole masses. We discuss the impact of these features in Section 6.

4 THE COVARIANCE MATRIX FOR POPULATION PARAMETERS

4.1 The Fisher information matrix

The Fisher matrix quantifies the amount of information that a set of observable random variables (in our case, the merger rate and chirp-mass distributions) carries about the parameters (in our case, the four tunable parameters described in Section 2) of a distribution that models these observables.

Specifically, the Fisher matrix F for a set of random variables $\mathcal{D}$ (the data) which are dependent on a set of parameters $\{\lambda\}$ is defined element-wise as

$$F_{ij} = -\left\langle \frac{\partial^2 \log[\mathcal{L}(\mathcal{D}|\{\lambda\})]}{\partial\lambda_i\,\partial\lambda_j}\right\rangle, \quad (15)$$

where $\mathcal{L}$ is the likelihood function, defined as the probability of acquiring the observed data $\mathcal{D}$ given the model parameters, and the angle brackets indicate an expectation over the data realization. We introduce the likelihood for our problem in the section below.

Under certain conditions, the inverse of the Fisher matrix gives a lower bound (the Crámer–Rao bound) on the covariance matrix for those dependent parameters (Vallisneri 2008); we discuss the regime of validity of the Fisher matrix inverse as an approximation to the covariance matrix in Section 5.2. The covariance matrix tells us how sensitive our data are to a change in the model parameters. We can also examine which combinations of dependent parameters are degenerate and which combinations yield the greatest information gain.[4]

The Fisher matrix quantifies the sensitivity of predicted observations to model parameters, and provides a bound on the accuracy of parameter inference. This approach assumes that the model is correct. The correctness of the model can be evaluated through other means. For example, model selection can be used to compare distinct models, whether these are different formation channels or different prescriptions for describing the physical processes of binary evolution (e.g. Mandel & O'Shaughnessy 2010; Stevenson et al. 2017b; Talbot & Thrane 2017; Vitale et al. 2017a; Zevin et al. 2017), or model-independent clustering can be used without reference to particular models (e.g. Mandel et al. 2015, 2017).

4.2 The COMPAS likelihood function

For this study we assume that we have a gravitational-wave catalogue of merging binary black holes, formed via the isolated binary evolution channel, and we focus on two observable characteristics of such a data set: the rate of detections and the distribution of chirp masses for the observed systems.

The likelihood function contains a term for each observational characteristic:

$$\log\mathcal{L}(\mathcal{D}|\{\lambda\}) = \log\mathcal{L}(N_\mathrm{obs}|\{\lambda\}, t_\mathrm{obs}) + \log\mathcal{L}(\{\mathcal{M}\}|\{\lambda\}). \quad (16)$$

The first term is the likelihood of observing binary black holes at a given rate. We assume that observable binary black holes coalesce in the Universe as a Poisson process with rate parameter μ, which is predicted by our population synthesis model, and the total number of observations N_obs, accumulated in a time t_obs. The Poisson likelihood is

$$\log\mathcal{L}(N_\mathrm{obs}|\{\lambda\}, t_\mathrm{obs}) = N_\mathrm{obs}\log(\mu t_\mathrm{obs}) - \mu t_\mathrm{obs} - \log(N_\mathrm{obs}!). \quad (17)$$

The second term is the likelihood of observing a given chirp-mass distribution. As described in Section 3, we have chosen to represent our chirp-mass distribution in bins. In this case the correct likelihood

[4]This is analogous to identifying the chirp mass as being the best measured combination of masses from gravitational-wave observations.

is a multinomial distribution (Stevenson et al. 2015)

$$\log \mathcal{L}\left(\{\mathcal{M}\}|\{\lambda\}\right) = \log(N_{\rm obs}!) + \sum_{k}^{K} \left[c_k \log(p_k) - \log(c_k!)\right], \quad (18)$$

where K is the number of chirp-mass bins, c_k is the number of observed systems falling into the k-th bin with $\sum_k c_k = N_{\rm obs}$, and p_k is the probability predicted by the model that a system falls into the k-th bin. Thus, μ and p_k are functions of the tunable model parameters λ, while c_k and $N_{\rm obs}$ are observables. Given the likelihood, we can now calculate the Fisher matrix.

4.3 Computing the Fisher matrix

In order to compute the Fisher matrix, we need to find the second derivatives of the likelihood with respect to the population parameters and average over the possible observations drawn according to the same likelihood distribution. First differentiating the total-rate log likelihood,

$$\begin{aligned}\frac{\partial^2 \log \mathcal{L}\left(N_{\rm obs}|\{\lambda\}\right)}{\partial\lambda_i\,\partial\lambda_j} &= \frac{\partial}{\partial\lambda_j}\left[\left(\frac{N_{\rm obs}}{\mu} - t_{\rm obs}\right)\frac{\partial\mu}{\partial\lambda_i}\right] \\ &= -\frac{N_{\rm obs}}{\mu^2}\frac{\partial\mu}{\partial\lambda_i}\frac{\partial\mu}{\partial\lambda_j} \\ &\quad + \left(\frac{N_{\rm obs}}{\mu} - t_{\rm obs}\right)\frac{\partial^2\mu}{\partial\lambda_i\,\partial\lambda_j}. \end{aligned} \quad (19)$$

Meanwhile, differentiating the chirp-mass portion of the log likelihood yields

$$\begin{aligned}\frac{\partial^2 \log \mathcal{L}\left(\{\mathcal{M}\}|\{\lambda\}\right)}{\partial\lambda_i\,\partial\lambda_j} &= \frac{\partial}{\partial\lambda_j}\left(\sum_k^K \frac{c_k}{p_k}\frac{\partial p_k}{\partial\lambda_i}\right) \\ &= \sum_k^K \left(-\frac{c_k}{p_k^2}\frac{\partial p_k}{\partial\lambda_i}\frac{\partial p_k}{\partial\lambda_j} + \frac{c_k}{p_k}\frac{\partial^2 p_k}{\partial\lambda_i\,\partial\lambda_j}\right). \end{aligned} \quad (20)$$

The expectation value of $N_{\rm obs}$ over this Poisson likelihood with rate parameter $\mu t_{\rm obs}$ is just $\langle N_{\rm obs}\rangle = \mu t_{\rm obs}$; similarly, $\langle c_k\rangle = \mu t_{\rm obs} p_k$. Therefore, the Fisher matrix is

$$F_{ij} = \mu t_{\rm obs}\left[\frac{1}{\mu^2}\frac{\partial\mu}{\partial\lambda_i}\frac{\partial\mu}{\partial\lambda_j} + \sum_k^K \frac{1}{p_k}\frac{\partial p_k}{\partial\lambda_i}\frac{\partial p_k}{\partial\lambda_j}\right], \quad (21)$$

where we used $\sum_k p_k = 1$ to eliminate the second term from equation (20). Crucially, this expression contains only first-order derivatives of the observables with respect to the population parameters. These derivatives can be readily and reliably estimated using population synthesis models, as described below.

4.4 Evaluating the first derivatives

We have shown in equation (21) that the Fisher matrix can be computed using just the first derivatives of the binned rates with respect to the population parameters. To compute derivatives, we simulated binary populations using a suite of variations to the population parameters discussed in Section 2.1. We used the same set of random seeds to the random number generator in COMPAS, so that for each variation the initial conditions (i.e. masses and separation) and random effects (i.e. kick directions) remain fixed. This allows us to accurately measure the derivatives by estimating the differential rather than absolute rates, reducing the uncertainty associated with a limited number of simulations.

We made six perturbations to the fiducial model for each population parameter (three negative and three positive). The perturbations were chosen to be sufficiently small that we could reliably estimate first derivatives numerically. A full list of the variations we used can be found in Table 1. For each of the quantities we are differentiating, we have a set of overconstrained simultaneous equations for the first and second derivatives according to the leading terms in the Taylor series, which we can write in a matrix form

$$\begin{pmatrix} f(\lambda+\Delta_1) - f(\lambda) \\ \vdots \\ f(\lambda+\Delta_6) - f(\lambda) \end{pmatrix} = \begin{pmatrix} \Delta_1 & \frac{1}{2}\Delta_1^2 \\ \vdots & \vdots \\ \Delta_6 & \frac{1}{2}\Delta_6^2 \end{pmatrix} \begin{pmatrix} \frac{\partial f(\lambda)}{\partial\lambda} \\ \frac{\partial^2 f(\lambda)}{\partial\lambda^2} \end{pmatrix}. \quad (22)$$

If we label the three terms in equation (22) as $\mathbf{y}$, $\mathbf{X}$, and β, respectively, then the maximum-likelihood solution for the derivatives $\hat{\beta}$ can be computed directly as (Anton & Rorres 2000, section 9.3)

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}. \quad (23)$$

We use this approach to compute all of the derivatives in equation (21) and combine them into an estimate of the Fisher matrix. The Fisher matrix can then be inverted to provide the Crámer–Rao lower bound on the covariance matrix of the astrophysical parameters evaluated at the COMPAS fiducial model.

Table 1. The 25 population-parameter variations used in this paper. The population parameters are described in Section 2.1: $\sigma_{\rm kick}$ is the dispersion parameter for a Maxwellian used to draw the magniutde of natal kicks from equation (1); $\alpha_{\rm CE}$ is the efficiency of common-envelope ejection from equation (4); $f_{\rm WR}$ is the multiplier for Wolf–Rayet wind mass-loss from equation (6), and $f_{\rm LBV}$ is the multiplier for LBV mass-loss described in equation (5). Our fiducial model appears in the top row. For each of these population parameter combinations we also varied metallicity. We used 12 different metallicities, which were evenly spaced in the log between $0.005\,Z_\odot$ and $Z_\odot$, where we use a solar metallicity $Z_\odot = 0.02$. We therefore had a total of 300 model variations. We simulated 1 197 989 binaries for each of these variations.

$\sigma_{\rm kick}$ (km s^{-1})	$\alpha_{\rm CE}$	$f_{\rm WR}$	$f_{\rm LBV}$
250.0	1.00	1.00	1.50
240.0	1.00	1.00	1.50
244.0	1.00	1.00	1.50
247.0	1.00	1.00	1.50
253.0	1.00	1.00	1.50
256.0	1.00	1.00	1.50
260.0	1.00	1.00	1.50
250.0	0.95	1.00	1.50
250.0	0.97	1.00	1.50
250.0	0.99	1.00	1.50
250.0	1.01	1.00	1.50
250.0	1.03	1.00	1.50
250.0	1.05	1.00	1.50
250.0	1.00	0.90	1.50
250.0	1.00	0.94	1.50
250.0	1.00	0.97	1.50
250.0	1.00	1.03	1.50
250.0	1.00	1.06	1.50
250.0	1.00	1.10	1.50
250.0	1.00	1.00	1.45
250.0	1.00	1.00	1.47
250.0	1.00	1.00	1.49
250.0	1.00	1.00	1.51
250.0	1.00	1.00	1.53
250.0	1.00	1.00	1.55

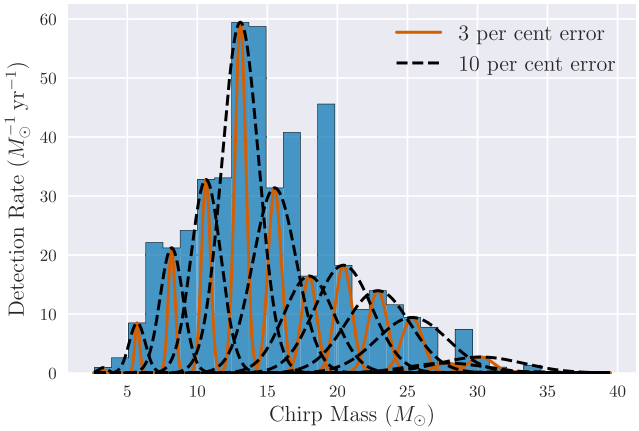

Figure 4. An illustration of how we include measurement errors in our analysis. A Gaussian is centred on each bin, with a standard deviation proportional to the value at the centre of that bin. That bin's counts are then distributed to other bins according to the fraction of that Gaussian falling in each bin.

4.5 Measurement uncertainty

The measurements of chirp masses will be subject to a certain amount of measurement uncertainty. We use a simplified treatment of this measurement uncertainty based on the methodology of Gair, Tang & Volonteri (2010, see their appendix A). We assume that the probability of finding a system in an incorrect bin is given by a Gaussian distribution about the centre of the correct bin into which the system was placed in the simulation.

Let f_i be the fraction of system predicted by the simulation to lie in the i-th bin, which is centred on chirp mass μ_i and has left and right edges at chirp masses μ_i^- and μ_i^+, respectively. Then the probability p_i of observing a system in the i-th bin is

$$p_i = \sum_{j=1}^{K} \frac{f_j}{\sqrt{2\pi\sigma_j^2}} \int_{\mu_i^-}^{\mu_i^+} \mathrm{d}x \, \exp\left[\frac{-(x-\mu_j)^2}{2\sigma_j^2}\right], \tag{24}$$

where σ_i is the standard deviation of the measurement in the i-th bin. In the limit of σ_i tending to zero, we recover perfect measurement accuracy, $p_i = f_i$. An illustration of this treatment of the measurement errors is presented in Fig. 4.

The chirp-mass measurement uncertainty depends strongly on the total mass of the source, with the most massive sources spending the fewest inspiral cycles in band, leading to the largest measurement uncertainty (e.g. Abbott et al. 2016a). It also scales inversely with the source SNR. Here, we crudely approximate this as a fixed fractional uncertainty on the chirp mass of 3 per cent (cf. Mandel et al. 2017; Vitale et al. 2017b). We therefore modify the binned rates according to equation (24), using a standard deviation $\sigma_i = 0.03\mu_i$.

This method of incorporating measurement errors is a simplification. The formally correct approach would be to incorporate them on a per-system basis, which would involve a modification of the likelihood function. Performing the analysis in this way would correctly account for correlations between bins, whereas in the simplified approach bins are modified independently, losing information, and slightly swelling the uncertainty.

4.6 Uncertainty quantification

The rate derivatives used to compute the Fisher matrix at the COMPAS fiducial model depend on the particular population realization used in the calculation. We quantify the impact of simulation realization noise, due to the finite number of simulated binaries, with bootstrapping. We recompute the Fisher matrix by re-creating data sets of the same size as the original simulated data set by drawing samples from it with replacement.

By repeating this process many times and observing the spread in the results, we can observe how much the point estimates change under different population realisations (different sets of binary initial conditions). Our full data set consists of 359 396 700 binary simulations, which consists of the same set of 1 197 989 ZAMS binaries evolved under each of 300 different model variations (the 25 population parameter combinations listed in Table 1, each simulated at the 12 different metallicities shown in Fig. 1). To generate one bootstrap sample Fisher matrix:

(i) We randomly choose 1 197 989 initial conditions, with replacement, from our original set of initial conditions.

(ii) For each of the 25 population parameter combinations in Table 1, we find the systems from the bootstrap initial conditions which become merging binary black holes, and calculate their total rate and chirp-mass distribution (taking into account cosmic history, selection effects, and measurement uncertainty).

(iii) We use equations (22) and (23) to compute the derivatives of the total rate and chirp-mass distribution bin heights, with respect to each population parameter.

(iv) We use these derivatives to compute the Fisher matrix, using equation (21).

We repeat the above steps 1500 times in order to express the uncertainty coming from the realization of the initial conditions, i.e. from the simulation statistical fluctuations. In principle, this model uncertainty could be overcome with more simulations, unlike the more fundamental uncertainties stemming from a finite number of observations and chirp-mass measurement uncertainty. We discuss the relative contributions of these sources of uncertainty in Section 6.

5 RESULTS AND DISCUSSION

Using the method described in Section 4 we computed the elements of the Crámer–Rao lower bound on the covariance matrix for the population parameters $\sigma_{\rm kick}$, $\alpha_{\rm CE}$, $f_{\rm LBV}$, and $f_{\rm WR}$. We computed simulation uncertainties on these elements by taking 1500 bootstrap samples from the 1 197 989 sets of initial conditions simulated for the binaries, specifically varying the metallicities, initial masses, and separations. Using these results, we are able to explore what can be learned about these population parameters using gravitational-wave observations of binary black holes. Results are presented for $N_{\rm obs} = 1000$ observations, a sufficiently large number to ensure the validity of our results; we discuss the effect of changing the number of observations in Section 5.2.

Fig. 5 shows the distribution of standard deviations of each of the population parameters. We see that it will be possible to measure $\alpha_{\rm CE}$, $f_{\rm LBV}$, and $f_{\rm WR}$ with fractional accuracies of $\sim$2 per cent after 1000 observations. We will be less sensitive to the value of $\sigma_{\rm kick}$. This is an expected result, since the natal kicks of black holes are reduced according to equation (2), and many of the more massive ones do not get a kick at all.

The fractional uncertainties on all of the parameters are quantities of order $N_{\rm obs}^{-1/2} \approx 0.03$ for $N_{\rm obs} = 1000$. Varying the parameters by their full dynamic range would change the rate by $\mathcal{O}(N_{\rm obs})$. For example, reducing $\alpha_{\rm CE}$ from 1 to 0 would make binary black hole

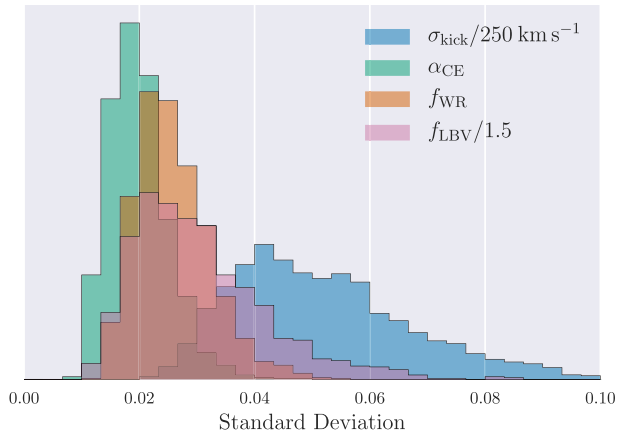

Figure 5. The inferred measurement accuracy for each of the four population parameters after observing 1000 systems, as estimated by taking the square root of the diagonal elements of the estimated covariance matrices for each of the 1500 bootstrapped sets. The histograms are normalised such that they all have the same area.

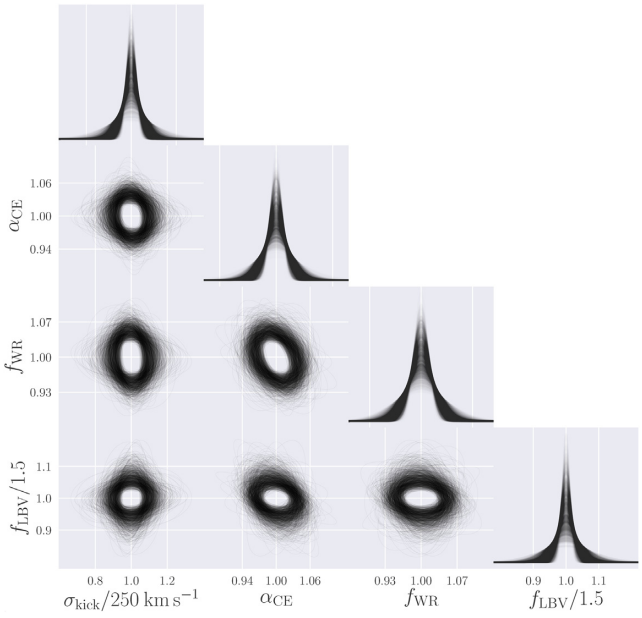

Figure 6. 1500 bootstrap samples of the marginalized univariate distributions and bivariate 90 per cent confidence intervals from the Crámer–Rao lower bound on the covariance matrix for the COMPAS population parameters. The univariate distributions are the Gaussian distributions corresponding to the standard deviations of Fig. 5, and have been normalized to have the same area.

formation through a common-envelope phase impossible, reducing the expected number of detections from $N_{\rm obs}$ to $\sim$0.

The measurement accuracy with which the tunable population parameters can be inferred using 1000 gravitational-wave observations can be alternatively interpreted from the perspective of model selection. For example, the median of the distribution for the standard deviation of $\alpha_{\rm CE}$ is $\sim$0.02. Therefore, if $\alpha_{\rm CE}$ different from the fiducial value by 6 per cent, the fiducial model could be ruled out with a confidence of $\sim 3\sigma \approx 99.7$ per cent.

We can examine the full multivariate normal behaviour of the population parameters. Fig. 6 shows marginalized univariate distributions and bivariate projections of the 90 per cent confidence

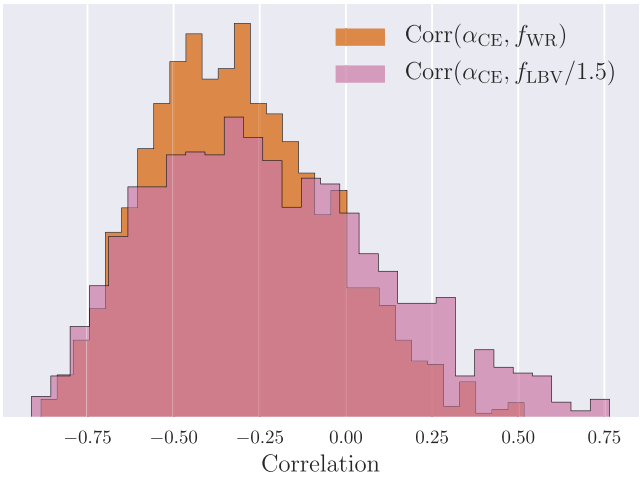

Figure 7. Distribution of correlations between $\alpha_{\rm CE}$ and each of $f_{\rm LBV}$ and $f_{\rm WR}$. The histograms have been normalized to have the same area.

interval for each of the bootstrap samples. This plot shows that most pairwise correlations between most population parameters are negligible. Fig. 7 shows the correlations between $\alpha_{\rm CE}$ and $f_{\rm WR}$, and between $\alpha_{\rm CE}$ and $f_{\rm LBV}$. Bootstrapping indicates an 88 per cent confidence that $\alpha_{\rm CE}$ and $f_{\rm WR}$ are anticorrelated. Increasing $\alpha_{\rm CE}$ increases the efficiency with which orbital energy is transferred into the common envelope. An increased efficiency means that there will be less tightening of the binary, so fewer systems will come sufficiently close together to merge within a Hubble time. Losing mass through winds widens the orbit, meaning that increasing the Wolf–Rayet wind mass-loss rate creates more systems which are too wide to merge within a Hubble time. Increased mass loss also results in the black holes being less massive, therefore increasing the time required for them to merge through gravitational-wave emission from a given initial separation (Peters 1964). These correlations mean that increasing (or decreasing) both $\alpha_{\rm CE}$ and $f_{\rm WR}$ would compound the effect on the rates, so their bivariate distribution (in Fig. 6) is narrower in this direction. Conversely, the effects of increasing one whilst decreasing the other would partially cancel out, and thus the bivariate distribution is wider in that direction. The confidence in the anticorrelation between $\alpha_{\rm CE}$ and $f_{\rm LBV}$ is only 76 per cent, and there is insufficient evidence for correlation between other parameter pairs.

5.1 Information from the total detection rate

To gain further insight into the correlations between the inferred parameters, we now consider what we could learn about the population parameters by considering only the total rate at which gravitational waves are observed. It is impossible to constrain the four-dimensional population parameter vector considered in this paper with a single observable, the binary black hole detection rate. In this case, all that can be learned about the population parameters is the value of some linear combination of them.

We construct a detection rate Fisher matrix, using only the total rate log likelihood of equation (17),

$$F_{ij}^{\rm RO} = \frac{t_{\rm obs}}{\mu}\frac{\partial\mu}{\partial\lambda_i}\frac{\partial\mu}{\partial\lambda_j}, \tag{25}$$

and perform an eigendecomposition. We expect to see that there is only one eigenvector whose eigenvalue is non-zero. We verified that this is true for all 1500 of our bootstrap samples, which provided a useful sanity check of our results.

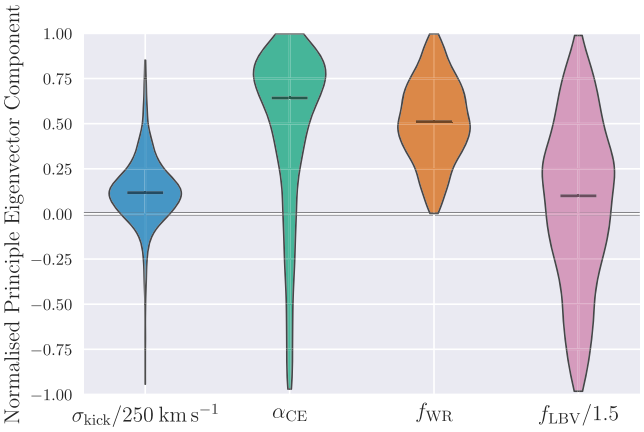

Figure 8. Violin plot showing components of the normalized principal eigenvector of the Fisher matrices calculated using only the total detection rate. The coloured regions give the bootstrapped distribution of the principle eigenvector direction, with medians marked in black.

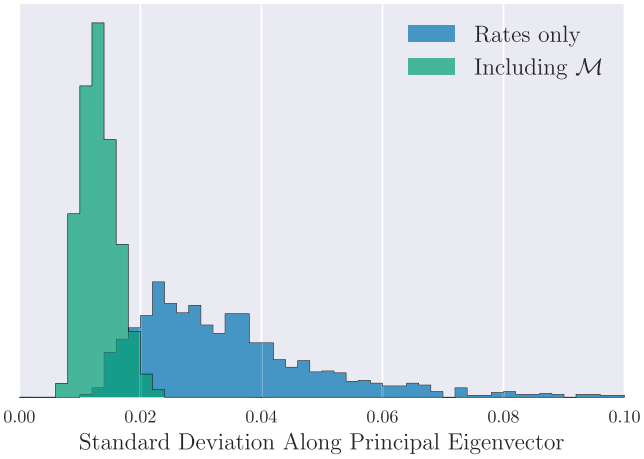

Figure 9. Distribution of the standard deviation of the particular linear combination of population parameters corresponding to the principal eigenvector of the total detection rate Fisher matrix. The measurement accuracy is computed using only information from the total rate (blue) and after including information from the chirp-mass distribution (green). The distributions come from considering all 1500 bootstrapped sets.

Next, by examining the eigenvector whose eigenvalue is non-zero, we can find the linear combination of population parameters to which we are sensitive. Fig. 8 shows a univariate representation of this direction (with its distribution from bootstrapping over simulations). The components of the vector parallel to $f_{\rm LBV}$ and $\sigma_{\rm kick}$ axes are broadly consistent with zero. Most of the information learned solely from the total detection rate is in the $\alpha_{\rm CE}$–$f_{\rm WR}$ plane. The fact that both values are simultaneously positive implies that they are correlated; this is the same correlation as was discussed at the beginning of this section.

Whilst we can only measure this specific combination of population parameters using only the total detection rate, we can constrain parameter combinations in the $\sim\alpha_{\rm CE} + f_{\rm WR}$ direction to within a few per cent from the total rate. Fig. 9 shows the standard deviation along the line defined by this combination of population parameters $a^{-1/2}$, where a is the principal eigenvalue. This can be interpreted in the same way as the standard deviations in Fig. 5, and matches the expected value of $\mathcal{O}(N_{\rm obs}^{-1/2})$. We see that if this combination of population parameters differed from our fiducial values by more than a few per cent, we would be able to confidently rule our model out after 1000 observations. However, we also see from Fig. 9 that including the chirp-mass distribution would significantly improve measurements of this parameter combination.

5.2 Number of observations

The expected number of observations only appears as a multiplicative term in equation (21), so that the standard deviations in Fig. 5 simply scale as $N_{\rm obs}^{-1/2}$. However, the results presented here are predicated on the assumption that the inverse of the Fisher information matrix is a good approximation to the covariance, and not just a lower bound. This in turn requires the likelihood to be approximately Gaussian, i.e. the linear single approximation (LSA; Vallisneri 2008) should hold. Only if the predicted parameter uncertainties are smaller than the neighbourhood in which the LSA is valid does the Fisher matrix provide a self-consistent estimate of the accuracy of parameter inference. This effectively sets a minimal threshold on the number of observations required for self-consistency in our estimates.

When computing the derivatives, as described in Section 4.4, we measure the terms in a Taylor expansion of an observable (binned) rate f as a function of the population parameter λ,

$$f(\lambda+\Delta) - f(\lambda) \approx \Delta f'(\lambda) + \frac{\Delta^2}{2} f''(\lambda). \tag{26}$$

In order to verify the validity of the LSA, we need to check that each f is indeed linear when Δ is of the same order as the computed standard deviations for the population parameters. We require that the linear term is dominant in the Taylor series, so that

$$f'(\lambda) \gg \frac{\Delta}{2} f''(\lambda). \tag{27}$$

We find $N_{\rm obs} = 1000$ to be a sufficient lower limit on the number of observations necessary to ensure the LSA is valid. At 1000 observations, the best measured combination of parameters is constrained at the per cent level, and this will continue to improve as we expand the catalogue of observations.

For smaller numbers of observations, the LSA will break down. The probability distribution for the model parameters may no longer be a multidimensional Gaussian so the Fisher matrix is likely to underestimate the inference uncertainty.

6 CONCLUSIONS

We have, for the first time, quantitatively analysed how accurately gravitational-wave observations of binary black hole mergers will constrain binary population synthesis models described by a multidimensional parametrization. When ground-based detectors have accumulated 1000 observations of merging binary black holes, we have shown that we will measure binary population synthesis model parameters with an accuracy of a few per cent. Equivalently, we will be able to distinguish models for which the population parameters differ only by a few per cent.

Our analysis accounts for three distinct sources of uncertainty in the inference of population parameters using gravitational-wave observations. The first is due to the finite number of observations. We show when the linear signal approximation holds (Section 5.2), the accuracy with which population parameters can be inferred scales

with the inverse square root of the number of observations. The second is the chirp-mass measurement uncertainty in individual observations. We only model this approximately (Section 4.5) but find that it is unlikely to be limiting factor in inference. The third source of uncertainty is simulation uncertainty: the accuracy in predicted detection rate and chirp-mass distribution is limited by the finite number of COMPAS simulations. This uncertainty, which we quantify with bootstrapping (Section 4.6), is only limited by computational cost, and be reduced indefinitely with more simulations or more efficient sampling (e.g. Andrews, Zezas & Fragos 2017).

There is, of course, potential systematic uncertainty in the binary evolution models themselves: for example, it is probable that the $\alpha_{\rm CE}$ parameter is not universal, as assumed here, but depends on the binary properties during the common-envelope phase. Model inference techniques such as those described here should be combined with model verification and with weakly modelled inference (e.g. Mandel et al. 2017).

We show the expected detection rate and chirp-mass distribution of merging binary black holes in Fig. 3. The sharp features in the chirp-mass distribution are due to only simulating systems at a small number (12) of metallicities, replacing the integral over metallicity in equation (9) with a discrete sum. Mass loss, particularly during the LBV phase, leads to a pile-up of black hole masses from the most massive stars at particular metallicity-dependent values. The subsequent discrete sum over metallicities overpopulates some bins in the chirp-mass distribution relative to neighbouring bins (cf. Dominik et al. 2013). This can impact our results, causing us to overstate the accuracy with which we will be able to measure population parameters. This issue can be addressed in the future by interpolating model predictions over metallicity (e.g. using Gaussian process emulators as described by Barrett et al. 2017), producing a smooth set of predictions.

Our primary intention with this paper was to introduce a methodology for evaluating the accuracy with which astrophysical model parameters can be estimated based on the rates and properties of observed transients. We considered a four-dimensional parameter space, but the number of dimensions is limited only by computational cost. It is also straightforward to add more terms than just the chirp-mass distribution to equation (16) in order to investigate other observable characteristics of binary populations such as mass ratios and spins (e.g. Stevenson et al. 2017b; Talbot & Thrane 2017; Zevin et al. 2017). Furthermore, this analysis can be used for other populations than observations of binary black hole mergers via gravitational-waves in this paper. Other observed populations, such as Galactic binary pulsars, X-ray binaries, Wolf–Rayet stars, short gamma-ray bursts, or luminous red novae (for a review, see De Marco & Izzard 2017), can provide orthogonal constraints on the parameters governing binary evolution (cf. Fig. 9). Over the coming decade, such measurements will help us to accurately determine the physics of massive binary evolution.

ACKNOWLEDGEMENTS

We thank Floor Broekgaarden and Jonathan Gair for useful comments and discussions, and the anonymous referee for constructive suggestions. We acknowledge the Science and Technology Facilities Council (STFC) who supported this work. AVG acknowledges support from the Consejo Nacional de Ciencia y Tecnologia (CONACYT). SS acknowledges support from the Australian Research Council Centre of Excellence for Gravitational Wave Discovery (OzGrav), through project number CE170100004.

REFERENCES

Aasi J. et al., 2015, Class. Quantum Gravity, 32, 074001
Abbott B. P. et al., 2016a, Phys. Rev. X, 6, 041015
Abbott B. P. et al., 2016b, Phys. Rev. Lett., 116, 241102
Abbott B. P. et al., 2016c, ApJ, 818, L22
Abbott B. P. et al., 2017a, Phys. Rev. Lett., 118, 221101
Abbott B. P. et al., 2017b, Phys. Rev. Lett., 119, 141101
Abbott B. P. et al., 2017c, ApJ, 851, L35
Abbott B. P. et al., 2018, Living Rev. Relativ., 21, 3
Abt H. A., 1983, ARA&A, 21, 343
Acernese F. et al., 2015, Class. Quantum Gravity, 32, 024001
Andrews J. J., Zezas A., Fragos T., 2017, AAS, preprint (arXiv:1710.11030)
Anton H., Rorres C., 2000, Elementary Linear Analysis, 8th edn. Wiley, New York
Babak S., Taracchini A., Buonanno A., 2017, Phys. Rev. D, 95, 024010
Barrett J. W., Mandel I., Neijssel C. J., Stevenson S., Vigna-Gómez A., 2017, in Brescia M., Djorgovski S. G., Feigelson E. D., Longo G., Cavuoti S., eds, Proc. IAU Symp. 325, Astroinformatics, Cambridge Univ. Press, Cambridge, p. 46
Belczynski K., Bulik T., Fryer C. L., Ruiter A., Valsecchi F., Vink J. S., Hurley J. R., 2010, ApJ, 714, 1217
Belczynski K., Holz D. E., Bulik T., O'Shaughnessy R., 2016, Nature, 534, 512
Bisnovatyi-Kogan G. S., 1993, Astron. Astrophys. Trans., 3, 287
Bray J. C., Eldridge J. J., 2016, MNRAS, 461, 3747
Bulik T., Belczyński K., 2003, ApJ, 589, L37
Bulik T., Belczyński K., Rudak B., 2004, A&A, 415, 407
Burrows A., Hayes J., 1996, Phys. Rev. Lett., 76, 352
Crowther P. A., 2007, ARA&A, 45, 177
Cutler C., Flanagan É. E., 1994, Phys. Rev. D, 49, 2658
de Kool M., 1990, ApJ, 358, 189
De Marco O., Izzard R. G., 2017, PASA, 34, e001
Dewi J. D. M., Tauris T. M., 2000, A&A, 360, 1043
Dominik M., Belczynski K., Fryer C., Holz D. E., Berti E., Bulik T., Mandel I., O'Shaughnessy R., 2012, ApJ, 759, 52
Dominik M., Belczynski K., Fryer C., Holz D. E., Berti E., Bulik T., Mandel I., O'Shaughnessy R., 2013, ApJ, 779, 72
Duchêne G., Kraus A., 2013, ARA&A, 51, 269
Eldridge J. J., Stanway E. R., 2016, MNRAS, 462, 3302
Eldridge J. J., Stanway E. R., Xiao L., McClelland L. A. S., Taylor G., Ng M., Greis S. M. L., Bray J. C., 2017, PASA, 34, e058
Finn L. S., 1992, Phys. Rev. D, 46, 5236
Finn L. S., Chernoff D. F., 1993, Phys. Rev. D, 47, 2198
Fryer C. L., Belczynski K., Wiktorowicz G., Dominik M., Kalogera V., Holz D. E., 2012, ApJ, 749, 91
Gair J. R., Tang C., Volonteri M., 2010, Phys. Rev. D, 81, 104014
Gerosa D., O'Shaughnessy R., Kesden M., Berti E., Sperhake U., 2014, Phys. Rev. D, 89, 124025
Giacobbo N., Mapelli M., Spera M., 2018, MNRAS, 474, 2959
Hamann W.-R., Koesterke L., 1998, A&A, 335, 1003
Hannam M., Schmidt P., Bohé A., Haegel L., Husa S., Ohme F., Pratten G., Pürrer M., 2014, Phys. Rev. Lett., 113, 151101
Hobbs G., Lorimer D. R., Lyne A. G., Kramer M., 2005a, MNRAS, 360, 974
Hobbs G., Lorimer D. R., Lyne A. G., Kramer M., 2005b, MNRAS, 360, 974
Holz D. E., Hughes S. A., 2005, ApJ, 629, 15
Humphreys R. M., Davidson K., 1994, PASP, 106, 1025
Hurley J. R., Pols O. R., Tout C. A., 2000, MNRAS, 315, 543
Husa S., Khan S., Hannam M., Pürrer M., Ohme F., Forteza X. J., Bohé A., 2016, Phys. Rev. D, 93, 044006
Ivanova N. et al., 2013, A&AR, 21, 59
Janka H.-T., 2013, MNRAS, 434, 1355
Janka H.-T., Mueller E., 1994, A&A, 290, 496
Khan S., Husa S., Hannam M., Ohme F., Pürrer M., Forteza X. J., Bohé A., 2016, Phys. Rev. D, 93, 044007
Krolak A., Schutz B. F., 1987, Gen. Relativ. Gravit., 19, 1163

Kroupa P., 2001, MNRAS, 322, 231
Kruckow M. U., Tauris T. M., Langer N., Szécsi D., Marchant P., Podsiadlowski P., 2016, A&A, 596, A58
Lamberts A., Garrison-Kimmel S., Clausen D. R., Hopkins P. F., 2016, MNRAS, 463, L31
Langer N., Norman C. A., 2006, ApJ, 638, L63
Madau P., Dickinson M., 2014, ARA&A, 52, 415
Mandel I., 2016, MNRAS, 456, 578
Mandel I., Farmer A., 2017, Nature, 547, 284
Mandel I., Farr W. M., Colonna A., Stevenson S., Tiňo P., Veitch J., 2017, MNRAS, 465, 3254
Mandel I., Haster C.-J., Dominik M., Belczynski K., 2015, MNRAS, 450, L85
Mandel I., O'Shaughnessy R., 2010, Class. Quantum Gravity, 27, 114007
Ma X., Hopkins P. F., Faucher-Giguère C.-A., Zolman N., Muratov A. L., Kereš D., Quataert E., 2016, MNRAS, 456, 2140
McClelland L. A. S., Eldridge J. J., 2016, MNRAS, 459, 1505
Mennekens N., Vanbeveren D., 2014, A&A, 564, A134
Miller M. C., 2016, Gen. Relativ. Gravit., 48, 95
Moe M., Di Stefano R., 2017, ApJS, 230, 15
Öpik E., 1924, Publications de L'Observatoire Astronomique de l'Universite de Tartu, 25, 6
Paczynski B., 1976, in Eggleton P., Mitton S., Whelan J., eds, Proc. IAU Symp. 73, Structure and Evolution of Close Binary Systems, Reidel, Dordrecht, p. 75
Pan Y., Buonanno A., Taracchini A., Kidder L. E., Mroué A. H., Pfeiffer H. P., Scheel M. A., Szilágyi B., 2014, Phys. Rev. D, 89, 084006
Peters P. C., 1964, Phys. Rev., 136, 1224
Planck Collaboration XIII, 2016, A&A, 594, A13
Postnov K. A., Yungelson L. R., 2014, Living Rev. Relativ., 17, 3
Renzo M., Ott C. D., Shore S. N., de Mink S. E., 2017, A&A, 603, A118
Repetto S., Igoshev A. P., Nelemans G., 2017, MNRAS, 467, 298
Sana H. et al., 2012, Science, 337, 444
Savaglio S. et al., 2005, ApJ, 635, 260
Smith N., 2017, Phil. Trans. R. Soc., 375, 20160268
Socrates A., Blaes O., Hungerford A., Fryer C. L., 2005, ApJ, 632, 531
Stevenson S., Berry C. P. L., Mandel I., 2017b, MNRAS, 471, 2801
Stevenson S., Ohme F., Fairhurst S., 2015, ApJ, 810, 58
Stevenson S., Vigna-Gómez A., Mandel I., Barrett J. W., Neijssel C. J., Perkins D., de Mink S. E., 2017a, Nature Communications, 8, 14906
Talbot C., Thrane E., 2017, Phys. Rev. D, 96, 023012
Tramper F., Sana H., de Koter A., 2016, ApJ, 833, 133
Vallisneri M., 2008, Phys. Rev. D, 77, 042001
Vangioni E., Olive K. A., Prestegard T., Silk J., Petitjean P., Mandic V., 2015, MNRAS, 447, 2575
Vink J. S., de Koter A., 2005, A&A, 442, 587
Vink J. S., de Koter A., Lamers H. J. G. L. M., 2001, A&A, 369, 574
Vitale S., Lynch R., Raymond V., Sturani R., Veitch J., Graff P., 2017b, Phys. Rev. D, 95, 064053
Vitale S., Lynch R., Sturani R., Graff P., 2017a, Class. Quantum Gravity, 34, 03LT01
Voss R., Tauris T. M., 2003, MNRAS, 342, 1169
Webbink R. F., 1984, ApJ, 277, 355
Wong T.-W., Valsecchi F., Ansari A., Fragos T., Glebbeek E., Kalogera V., McClintock J., 2014, ApJ, 790, 119
Woosley S. E., 1987, in Helfand D. J., Huang J.-H., eds, Proc. IAU Symp. 125, The Origin and Evolution of Neutron Stars, D. Reidel Publ. Co., Dordrecht, p. 255
Yoon S.-C., 2017, MNRAS, 470, 3970
Zevin M., Pankow C., Rodriguez C. L., Sampson L., Chase E., Kalogera V., Rasio F. A., 2017, ApJ, 846, 82

This paper has been typeset from a TEX/LATEX file prepared by the author.